

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 09:53:54 ; Search time 88.9 seconds  
(without alignments)  
1015.782 Million cell updates/sec

Title: US-09-701-711-4  
Perfect score: 4202  
Sequence: 1 MRNSYFKQFSAMTMAVM.....LNKKONDQDTVQFQIGSVF 813

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues  
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_032802.\*  
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4202	100.0	813	21	AAV44390 M. catarrhalis BAS
2	4189	99.7	813	21	AAV44391 M. catarrhalis (AT
3	1123.5	26.7	792	22	AAU03958 Neisseria meningit
4	1123.5	26.7	792	22	AAU03958 Amino acid sequenc
5	1123.5	26.7	792	22	AAU03957 Neisseria gonorrhoe
6	1115	26.5	797	22	AAU03957 Neisseria meningit
7	1115	26.5	797	22	AAU03957 Amino acid sequenc
8	1115	26.5	797	22	AAU03957 Neisseria meningit
9	1111	26.4	797	22	AAU03959 Neisseria gonorrhoe
10	1111	26.4	797	22	AAU04451 Neisseria meningit
11	1111	26.4	797	22	AAU04451 Amino acid sequenc

12	1111	26.4	797	22	AAU03958
13	1109	26.4	797	21	AAV44390
14	1104.5	26.3	792	21	AAV44391
15	1052.5	25.0	797	15	AAU03958
16	1051.5	25.0	797	15	AAU03957
17	1050.5	25.0	797	15	AAU03956
18	1047.5	24.9	793	15	AAU03958
19	1042	24.8	792	15	AAU03957
20	901	21.4	639	22	ABG17654
21	569	13.5	896	22	ABG17654
22	434	10.3	916	22	ABG17654
23	433	10.3	847	19	AAW71477
24	429	10.2	925	18	AAW57229
25	429	10.2	925	20	AAW17183
26	395	9.4	795	20	AAW34896
27	393	9.4	787	21	AAW69362
28	386.5	9.2	797	20	AAW69362
29	379.5	9.0	899	20	AAW34369
30	378.5	9.0	891	20	AAW34495
31	356.5	8.5	837	20	AAW52774
32	348.5	8.3	812	20	AAW52775
33	344.5	8.2	614	18	AAW55643
34	311	7.4	1249	22	ABG21666
35	256.5	6.1	578	21	AAW5820
36	249.5	5.9	576	21	AAW5821
37	198.5	4.7	469	21	AAW44453
38	197.5	4.7	469	21	AAW44550
39	197.5	4.7	469	22	AAW2615
40	180.5	4.3	624	21	AAW32162
41	180.5	4.3	643	21	AAW32161
42	180.5	4.3	818	21	AAW32160
43	176.5	4.2	919	21	AAW5987
44	176.5	4.2	1092	19	AAW41602
45	175.5	4.2	889	21	AAW5988

ALIGNMENTS

RESULT 1  
AAV44390  
ID AAV44390 standard; Protein; 813 AA.  
AC AAV44390;  
XX  
XX  
DT 14-MAR-2000 (first entry)  
XX  
DE M. catarrhalis BAS027 polypeptide #1.  
XX  
KW BAS027; OMP85; outer membrane protein; otitis media; treatment;  
KW diagnosis; bacterial infection.  
XX  
XX Moraxella catarrhalis.  
XX  
XX WO9963093-A2.  
XX  
PD 09-DEC-1999.  
XX  
PF 31-MAY-1999; 99WO-EP03822.  
XX  
PR 03-JUN-1998; 98GB-0011945.  
PR 08-MAR-1999; 99GB-0005304.  
XX  
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
XX  
XX Vinals-Bassols C;  
XX  
XX WPI; 2000-105700/09.  
XX  
XX N-PSDB; RAZ29550.  
XX  
XX Novel BAS027 polynucleotide and polypeptides from Moraxella  
XX  
XX Moraxella catarrhalis useful for treating M. catarrhalis infection such as otitis  
XX media

XX PS  
XX PS  
XX PS  
CC The present sequence is BASB027 polypeptide, which shows significant  
CC homology to Neisseria meningitidis OMP85 outer membrane protein. It is  
CC encoded by DNA obtained from chromosomal DNA library of Moraxella  
CC catarrhalis strain Mc2931 (ATCC 43617). BASB027 polynucleotide and  
CC polypeptide can be used for diagnosis and staging of disease, determining  
CC susceptibility to a disease and to prepare medicaments for treating M.  
CC catarrhalis infections, especially otitis media. The BASB027 DNA can be  
CC used as probe for screening of genetic mutations, serotype, taxonomic  
CC classification or identification. BASB027 agonists, antagonists and  
CC antibodies may be used to prevent and/or treat bacterial infections.

XX Sequence 813 AA;

Query Match 100.0%; Score 4202; DB 21; Length 813;  
Best Local Similarity 100.0%; Pred. No. 5.7e-275;  
Matches 813; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MRSYFKGFOVSAMTAMVMVMSTHAQAADFMANDDITITGLQRTVIESLQSVLPRLQGV 60  
Db 1 mrsyfkfgvmsamtmvmvmsthaqaadfmnddititglqvieslqsvlpfirlgqv 60

Qy 61 VSENLADGVKALYATGNFSDVQVYHOGRIIYQVTERPLIAEINFEGNRLIPKGLQEG 120  
Db 61 vseqladgvkalyatgnfsvdvqvyhogrliiyqvtterpliaeinfegnrlipkeglqeg 120

Qy 121 LKNAGLAVGQPLKQATVOMIETELTNOYISQGYNTETVTKOTMLDGNRVKLDMTFAEGK 180  
Db 121 lknaglavgqplkqatvqmieteltnoyisqgyntetvktotmldgnrvklmtfaegk 180

Qy 181 PARVVDINIGNQHFSDADLDVLAIDKINKPLSKADRYTOEKLVTSLLENRAKYLNAG 240  
Db 181 parvvdinignqhfdsadldvllaikdnkplskadrytoeklvtslenlrakyl nag 240

Qy 241 FVFEIKDALKINEDKNRIFVEISLHEGEYRFQGTQFLGNLYTQAELEALLKFAEE 300  
Db 241 fvfeikdaklinedknri fveislhegeyrfqgtqflgnlytqaeleallkfaee 300

Qy 301 GFQSMLEQTTNNSTKFGDGYAAQIRPVTRINDESRVTDVEYIDPVHPVYVRRNF 360  
Db 301 gfsqsmleqttnnstkfgdgyaaqirpvtrindesrvtdveyidpvhpvyvrrnf 360

Qy 361 TGNETQDEVLRREMRQLEGALASQKIQLSRARLMRTGFFKHVTVTRPVNSPDQYDV 420  
Db 361 tgnftqdevlrremrqlegalasnkqiklsrarlmrtgffkhvtrpvnsdpdqv 420

Qy 421 NFVVEEQSGSSTTAAGYSQSGGVTFQFDVQNNPMGTGKHVNASFSSRETVYSLGMT 480  
Db 421 nfveeqsgsstaagysqsggvtfqfdvqnnpmgtgkhvnasfssretrevyslgmt 480

Qy 481 NPYFTVNGVQSLSGYRRTKYDNKNISNYLDSYGGSLSYGYPIDENORISFGLNADNT 540  
Db 481 npyftvngvqs slsgyr rtkydnknisnyl dsgslsygypidenorisfglnadnt 540

Qy 541 KLHGGREMGISNVKQLMADGGKIQVNDNGIPDFKHDTYTYNAILGNWYSSLDPRVPFTQG 600  
Db 541 klhggremgisnvkqlmadgk iqvndngipdfkhdt yty nailgnw yssldprvpftqg 600

Qy 601 MSHSVDLTVGDKTHQKVYVQGNIRYRFTKSVLRGAKLYGNNLPFENFYAGGYGS 660  
Db 601 mshsvdltv gdkthqkv yvqgniry rftksvl rgakly gnnlpfenfyaggygs 660

Qy 661 VRGYDQSSILGRPSQAYLTARQQOFTLGEVVGGNALATFGSELILPLPKGDWIDQVRPV 720  
Db 661 vrgydqssilgrpsqayltarqqoftlgevv ggnalatfgselilplpkgdw idqvrpv 720

Qy 721 IFIEGGQVFTDGTGMDKQFTIDLTQFKDPQATAEQNAKANRPLLTQDKOLRYSAGVATWY 780  
Db 721 ifieggqvftdgtgmdkqftidltqfkdpqataeqnak anrplltqdkolrysagvatwy 780

Qy 781 TPIGPLSISYAKPLNKKKQNDQDTVQFOIGSVF 813  
Db 781 tpigplsisyakplnkkqndqdtvqf igsvf 813

RESULT 2  
ID AAY44391 standard; Protein; 813 AA.  
XX AAY44391;  
XX AAY44391;  
XX 14-MAR-2000 (first entry)  
XX M. catarrhalis (ATCC 43617) BASB027 polypeptide #2.  
XX BASB027; OMP85; outer membrane protein; otitis media; treatment;  
XX diagnosis; bacterial infection.  
XX Moraxella catarrhalis.  
XX WO9963093-A2.  
XX 09-DEC-1999.  
XX 31-MAY-1999; 99WO-EP03822.  
XX 03-JUN-1998; 98GB-0011945.  
XX 08-MAR-1999; 99GB-0005304.  
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
XX Vinals-Bassols C;  
XX WPI; 2000-105700/09.  
XX N-PSDB; AAZ29551.  
XX Novel BASB027 polynucleotide and polypeptides from Moraxella  
XX catarrhalis useful for treating M. catarrhalis infection such as otitis  
XX media  
XX Claim 1; Page 102-104; 109pp; English.  
XX The present sequence is BASB027 polypeptide, which shows significant  
XX homology to Neisseria meningitidis OMP85 outer membrane protein. It is  
XX encoded by DNA obtained from chromosomal DNA library of Moraxella  
XX catarrhalis strain Mc2931 (ATCC 43617). BASB027 polynucleotide and  
XX polypeptide can be used for diagnosis and staging of disease, determining  
XX susceptibility to a disease and to prepare medicaments for treating M.  
XX catarrhalis infections, especially otitis media. The BASB027 DNA can be  
XX used as probe for screening of genetic mutations, serotype, taxonomic  
XX classification or identification. BASB027 agonists, antagonists and  
XX antibodies may be used to prevent and/or treat bacterial infections.

XX Sequence 813 AA;

Query Match 99.7%; Score 4189; DB 21; Length 813;  
Best Local Similarity 99.8%; Pred. No. 4.3e-274;  
Matches 811; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MRSYFKGFOVSAMTAMVMVMSTHAQAADFMANDDITITGLQRTVIESLQSVLPRLQGV 60  
Db 1 mrsyfkfgvmsamtmvmvmsthaqaadfmnddititglqvieslqsvlpfirlgqv 60

Qy 61 VSENLADGVKALYATGNFSDVQVYHOGRIIYQVTERPLIAEINFEGNRLIPKGLQEG 120  
Db 61 vseqladgvkalyatgnfsvdvqvyhogrliiyqvtterpliaeinfegnrlipkeglqeg 120

Qy 121 LKNAGLAVGQPLKQATVOMIETELTNOYISQGYNTETVTKOTMLDGNRVKLDMTFAEGK 180  
Db 121 lknaglavgqplkqatvqmieteltnoyisqgyntetvktotmldgnrvklmtfaegk 180

Qy 181 PARVVDINIGNQHFSDADLDVLAIDKINKPLSKADRYTOEKLVTSLLENRAKYLNAG 240



```

Db 181 parvvdinlignghfsdadliavlakdnklnplskadrytqeklvtslenlrakylmag 240
Qy 241 fvrfeikdaklnedknrfveislhegeoyrfqotplgnltvtoaeleallkfkake 300
Db 241 fvrfeikdaklnedknrfveislhegeoyrfqotplgnltvtoaeleallkfkake 300
Qy 301 gfsqamleqtnnistkfgdgyyayirpvttrinidesrtvveyidvpvhvrrinf 360
Db 301 gfsqamleqtnnistkfgdgyyayirpvttrinidesrtvveyidvpvhvrrinf 360
Qy 361 tgnfktqdevlrrmqlegalsnqkqlsrarlmrtgffkhvtdvtrvpnsdpqdv 420
Db 361 tgnfktqdevlrrmqlegalsnqkqlsrarlmrtgffkhvtdvtrvpnsdpqdv 420
Qy 421 nfvveeqpsgstiaagysqsgvtfqdvsonnmfmgtkhvnasfsrsetrevyslgmt 480
Db 421 nfvveeqpsgstiaagysqsgvtfqdvsonnmfmgtkhvnasfsrsetrevyslgmt 480
Qy 481 nfyftvngvsqslsgyrrtkydnknisnyldsygsglsygyypidenqrifsglnadnt 540
Db 481 nfyftvngvsqslsgyrrtkydnknisnyldsygsglsygyypidenqrifsglnadnt 540
Qy 541 klhggrfmgisnkvqlmadgkqivdngngipdfkhdtytnailgwnyslslrprvftqg 600
Db 541 klhggrfmgisnkvqlmadgkqivdngngipdfkhdtytnailgwnyslslrprvftqg 600
Qy 601 mshsvdltvfgdkthqkvvyogntyrpfikksvlrgvaklgcggnlpyenfyaggygs 660
Db 601 mshsvdltvfgdkthqkvvyogntyrpfikksvlrgvaklgcggnlpyenfyaggygs 660
Qy 661 vrgyqsslgprsqayltarrgqotlgevgngnalatfgselilplpfgkdwidqvrvp 720
Db 661 vrgyqsslgprsqayltarrgqotlgevgngnalatfgselilplpfgkdwidqvrvp 720
Qy 721 fiegqgvfdtgmokqotidltqfkdqataeqnakaanrpiltdqklyrsagvgatwy 780
Db 721 fiegqgvfdtgmokqotidltqfkdqataeqnakaanrpiltdqklyrsagvgatwy 780
Qy 781 tpigplslyakplnknkondtvtvqfoigsvf 813
Db 781 tpigplslyakplnknkondtvtvqfoigsvf 813

RESULT 3
AAU03958
ID AAU03958 standard; Protein; 792 AA.
XX
AC AAU03958;
XX
DT 23-OCT-2001 (first entry)
XX
DE Neisseria meningitidis serogroup A antigenic protein #1.
XX
KW Serogroup B antigen; pharynx; meningitis; septicaemia; mammalian cell;
KW bacterial infection; baculovirus; yeast.
XX
OS Neisseria meningitidis.
XX
FH Key Location/Qualifiers
FT Peptide 1..21
FT /note= "Signal peptide"
FT Protein 22..792
FT /note= "Mature N. meningitidis serogroup A antigen"
XX
PN WO200138350-A2.
XX
PD 31-MAY-2001.
XX
PF 28-NOV-2000; 2000WO-IB01851.
XX
PR 29-NOV-1999; 99GB-0028197.
PR 09-MAR-2000; 2000GB-0005698.

```

```

XX (CHIR-) CHIRON SPA.
PA (STAT-) STATENS INST FOLKEHELSE.
XX
PI Giuliani MM, Pizsa M, Rappuoli R, Holst J;
XX
WPI; 2001-381289/40.
DR N-PSDB; AA507278.
XX
Novel 85 kDa antigen from Neisseria meningitidis and Neisseria
gonorrhoeae, useful in the manufacture of a medicament for treating and
preventing Neisserial bacteria infection -
XX
Claim 1; Page 66-68; 92pp; English.
XX
The sequence represents a Neisseria meningitidis serogroup A 85 kDa
antigenic protein. Neisseria meningitidis colonises the pharynx, causing
meningitis and, occasionally, septicaemia in the absence of meningitis.
This antigenic protein is useful in the manufacture of a medicament for
treating or preventing infection due to Neisseria bacteria, such as
meningitis and septicaemia. It is also useful as a diagnostic reagent for
detecting the presence of Neisseria bacteria or antibodies raised against
Neisseria, and as a reagent for raising the antibodies. The Neisserial
nucleotide sequences can be expressed in a variety of different
expression systems, for example, mammalian cells, baculoviruses, plants,
bacteria and yeast.
CC Note: There are two versions of this sequence displayed in the
CC specification (see AAU04451).
XX
SQ Sequence 792 AA;
XX
Query Match 26.7%; Score 1123.5; DB 22; Length 792;
Best Local Similarity 32.6%; Pred. No. 2.3e-67;
Matches 273; Conservative 153; Mismatches 332; Indels 79; Gaps 19;
Qy 10 QVSAMTAYMYMSTHAQAADFMANDITITGLQRTVIESLQSLVPLRGLQGVVSENADLG 69
Db 2 kklqasalmmlgslafadftiqdrveglrtqepstvfnylvkvgdyndthgsai 61
Qy 70 VKALYATGNFSDVOYVHQEGRIYQVTRPLTAETAEINFEGRNLPKLEGLKAGLAVG 129
Db 62 kkslyatgffdrvretadgqllltvierptgslntgkmlqndaklnesflags 121
Qy 130 QPLKQATQVMIETELTNOYISOGVYNTETITVKOTMLDGNRVKLDMTFAEGKPARVVDINI 189
Db 122 qyfnqatinqavaglkelylgrgklniqitpkvklarnrvdidiididegksakidief 181
Qy 190 IGQHFSDADLDVLAIKDNKI-NPLSKADRYTQEKLVTSLENLRKAYLNAGVFRFEIKD 248
Db 182 egnyvsdrklmrqmslttegiwltlrsdrfdrqfaadmekvtfyqngydfriid 241
Qy 249 AKLNINEDKNRIFVEISLHEGEQYRFGQTFGLNLT-YTOAELEALLKFKAEFGSQAML 307
Db 242 tdiqtnedktrtkitvhhegrfrwkgvsiiegdtnevpkaeleklitmkpgkwyerqgm 301
Qy 308 EOTTNNISTKFGDDGYVYQAIQVTRINDESRTVDVEYIDVPVHPVYVRRINFTGFKTQ 367
Db 302 tavlgeiqnrmgsagyseisvqplpnagtktdvflhiepprklyvneihitgnnkr 361
Qy 368 DEVLRRMRQLEGALASNOKIQLSRARLMRTGFFKHVTVDTRPVPNSPQVDVNFVVEEQ 427
Db 362 devvrrelrqmesapydtsklrskervellgyfdnvqfdavplagtpdkvdlmnlalter 421
Qy 428 PGSSTIAAGYSQSGGVTFQFDVSONNFMGTGKHVNASFSSRSETREYVSLGWTNPFYTVN 487
Db 422 stgsidlsagwvqdtglvmsagvsqdnlfgtksaalrasrskttingslsfcdpyftad 481
Qy 488 GVSQSLSGY--YRKT---KYDNKNISNYLDYSGSLSYGYPIDENQRIISFGLNADNTKL 542
Db 482 gvs---lgydlygkafdrprkastsvkqytkttaggvgvmgipvteydrvnfglaehltv 538
Qy 543 HG-----GRFMGISNVKQLMADGGKIQVDNNGIPDFKHDTYTTYNAILGWN 587

```

Db 539 ntynkapyadfirkygtdg-----adg-----sfkg--llykgtvgwg 577  
 Qy 588 YSSLDPRVPTQGMSSHVDLTVGF-GDK-THQKVYVQGNIVRPPIKKSVLRL-----GYAK 640  
 Db 578 rnktdsaswptrgyltgvnaeialpgsklqyysathnqtffplsktftlmllggevagiag 637  
 Qy 641 LGYG--NNLPFENFYAGGYSVGRYDSSLSGPRSQAYLTARRQQQTTLGGEV--GGNAL 696  
 Db 638 -gygrtkeipfenfyggglgsrvgyesgtlpgk-----vydeygekiisygnk 686  
 Qy 697 ATFGSELILPLPFKGDWIDQVRPVIFFTEGGQVFTTGMKQTDIDLTQFKDPQATAEQNAK 756  
 Db 687 anvsaellfmpgagk-artvrlsfadagsv-----dgrtytaeangnknksyvenah 740  
 Qy 757 AANRPLLTQDKOLRYSAGVGATWTPICPLSISYAKPLNKKONDQTDVTFQIGSVF 813  
 Db 741 ks-----tftnelrysagavtwsplgpmkfsyayplkkpdeiqrfqfqlgttf 792  
 RESULT 4  
 AAB84745  
 ID AAB84745 standard; Protein; 792 AA.  
 AC AAB84745;  
 XX  
 DT 17-SEP-2001 (first entry)  
 XX  
 DE Amino acid sequence of a Neisseria gonorrhoeae protein.  
 XX  
 KW Serogroup B protein; outer membrane protein; Neisserial infection;  
 XX vaccine.  
 OS Neisseria gonorrhoeae.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /note= "signal peptide"  
 FT Protein 22..792  
 FT /note= "mature protein"  
 XX  
 PN WO200152885-A1.  
 XX  
 PD 26-JUL-2001.  
 XX  
 XX 17-JAN-2001; 2001WO-IB00166.  
 XX  
 XX 17-JAN-2000; 2000GB-0001067.  
 PR 09-MAR-2000; 2000GB-0005699.  
 XX  
 XX (CHIR-) CHIRON SPA.  
 PA  
 XX Pizza M, Rappuoli R, Giuliani M;  
 PI  
 XX WPI; 2001-451895/48.  
 DR N-PSDB; AAH42129.  
 XX  
 XX Composition for treating or preventing infection to, detecting, or for  
 PT raising antibodies against Neisserial bacteria, comprises an N.  
 PT meningitidis serogroup B outer membrane preparation and an immunogenic  
 PT component -  
 XX  
 XX Disclosure; Page 65-67; 83pp; English.  
 PS  
 XX The present sequence represents a Neisseria gonorrhoeae protein. The  
 CC protein is used to produce the compositions of the invention. The  
 CC specification describes a composition, comprising a Neisseria  
 CC meningitidis serogroup B outer membrane preparation and an immunogenic  
 CC component. The immunogenic component is protein disclosed in WO99/57280,  
 CC WO99/36544, WO99/24578, WO99/66791, WO97/28273, WO96/29412, WO95/03413,  
 CC WO99/31132, WO99/58683, WO99/55873, and/or N. meningitidis protein PorA,  
 CC TbpA, TbpB, PilC, Opa, or Omp85. The composition is used for making  
 CC a medicament for treating or preventing infection due to Neisserial

CC bacteria; a diagnostic reagent for detecting the presence of Neisserial  
 CC bacteria or of antibodies raised against Neisserial bacteria; and/or  
 CC a reagent which can raise antibodies against Neisserial bacteria. It may  
 CC also be used as a vaccine.  
 XX  
 SQ Sequence 792 AA;  
 Query Match 26.7%; Score 1123.5; DB 22; Length 792;  
 Best Local Similarity 32.6%; Pred. No. 2.3e-67;  
 Matches 273; Conservative 153; Mismatches 332; Indels 79; Gaps 19;  
 Qy 10 QVSAMTWAVMMVMTTHAQADFMANDTITGLQRTVIESLOSVLPRFGRQVVSNNQADG 69  
 Db 2 kklqiasalmmglspiafaftidqirveglrtqstvfnyipvkvgdydthgsai 61  
 Qy 70 VKALYATGNESDVOYHQEGRIIYQVTERPLIAEINFEENRLIPKEGLEGLKKNAGLAVG 129  
 Db 62 klsyatgfdvrvetadgqllltvierptgslntgkmlqndalkklesfglaqs 121  
 Qy 130 QPLKOATVQMIETELTNQYISQGYNTETITVKOTMLDGNRVKLDMTFAEGKPARVVDINI 189  
 Db 122 qyfnqatinqavaglkkeylgrgklniqitpkvcklarnrvdiditidegksakitdief 181  
 Qy 190 IGNOHFSADLIDVLAIKDKNI-NPLSKADRYTOEKLVTSLNLRKAYLNAGFVRFEKD 248  
 Db 182 egngvysdrklmrqmsltteggitwlttrsdrrfkqdaqmekvtfdygnngydfdrild 241  
 Qy 249 AKLNINEDKNRIFVEISLHEGEQYRFGOTQFLGNT-YTQAELEALLKFAKEGFSQAML 307  
 Db 242 tdiqtnedktrqtkitvhiegrrfwgkvsiegdtnevpkaelekillmkpgkwyergqm 301  
 Qy 308 EQTTNNISTKFGDDGYYYAQIRPVTRINDESRVTDVEYIIDPVHPVYVRRINFTGNFKTQ 367  
 Db 302 tavlgelqnrmsagyselssvqplpnagktvdfvlhiepgkryvneihiggnktr 361  
 Qy 368 DEVLREMRQLEGALASNOKIQLSRARLMRTGFKKHVTVTRPVNSDPDQVDVNFVEEQ 427  
 Db 362 devvrrelrimesapydtsklqrskervellgyfdnvqfdavpagnagtdpkvdlmslter 421  
 Qy 428 PSGSSTIAAGYSQSGGVTFQFDVSONNFMGTGKHVNASFSRSETRREVYSLGTMTPYFVFN 487  
 Db 422 stgsldisagvwvqdtglvmsagvsqdnlfgtgksaalrasrsktllngslsftqpyftad 481  
 Qy 488 GVSQSLSGY--YRKT---KYDNKNISYVLDSSYGSLSYGPIDENQRISEFLNADNTKL 542  
 Db 482 gvs--lgydlygkafdpkrastsvkqyktttagggvrmgipvtteydrvnfglaeahitv 538  
 Qy 543 HG-----GRFMGISNVKQLMADGGKIQVDNNGIIPDFKHDYTYNAILGNW 587  
 Db 539 ntynkapyadfirkygtdg-----adg-----sfkg--llykgtvgwg 577  
 Qy 588 YSSLDPRVPTQGMSSHVDLTVGF-GDK-THQKVYVQGNIVRPPIKKSVLRL-----GYAK 640  
 Db 578 rnktdsaswptrgyltgvnaeialpgsklqyysathnqtffplsktftlmllggevagiag 637  
 Qy 641 LGYG--NNLPFENFYAGGYSVGRYDSSLSGPRSQAYLTARRQQQTTLGGEV--GGNAL 696  
 Db 638 -gygrtkeipfenfyggglgsrvgyesgtlpgk-----vydeygekiisygnk 686  
 Qy 697 ATFGSELILPLPFKGDWIDQVRPVIFFTEGGQVFTTGMKQTDIDLTQFKDPQATAEQNAK 756  
 Db 687 anvsaellfmpgagk-artvrlsfadagsv-----dgrtytaeangnknksyvenah 740  
 Qy 757 AANRPLLTQDKOLRYSAGVGATWTPICPLSISYAKPLNKKONDQTDVTFQIGSVF 813  
 Db 741 ks-----tftnelrysagavtwsplgpmkfsyayplkkpdeiqrfqfqlgttf 792  
 RESULT 5  
 AAB23786  
 ID AAB23786 standard; Protein; 792 AA.  
 XX

[illegible]



Query Match	26.5%;	Score 1115;	DB 22;	Length 797;
Best Local Similarity	32.5%;	Pred. No. 8.8e-67;		
Matches 266;	Conservative 154;	Mismatches 361;	Indels 38;	Gaps 14;
QY	10 QVSAMTMAVMVMNSTHAQAADFMANDITITGLQRTVIESLQSVLPFRLGQVSVENQLADG	69		
Db	2 ktkqiasalmmlgisplaladftiqdirvegqrtepstvfnylpkvvgdyndthgsai	61		
QY	70 VKALYATGNFSDVOVYHQEGRIIYQVTERPLIAEINFEGRNRLIPKEGLQGLKNAGLAVG	129		
Db	62 ikslyatgfdvrvetadgqllltvterptgslntgskmlqndaknklesfglags	121		
QY	130 QPLKQATVQMIETELTNOYISOGYNTETVKTMLDGNRVKLDMTFAEGKPARVVDINI	189		
Db	122 qyfnqatlnqavaglkceylgrgklniqitpkvtklarnrvdiditidegksakitdef	181		
QY	190 IGNQHFSADLIDVLAIKDKNI-NPLSKADRYTQEKLVTSLENLRKAKYLVFRFFIKD	248		
Db	182 egngvysdrklmrqmsltggtwtlrsnqfneqfagdmekvtfdygnngyfdfrild	241		
QY	190 IGNQHFSADLIDVLAIKDKNI-NPLSKADRYTQEKLVTSLENLRKAKYLVFRFFIKD	248		
Db	182 egngvysdrklmrqmsltggtwtlrsnqfneqfagdmekvtfdygnngyfdfrild	241		
QY	249 AKLNINEDKNRIFVEISLHGEQYRFGQOTFLGNLT-YTQAELEALLKFAEGFSQAML	307		
Db	242 tdiqtnektktikitvhggrfwgksiegdtnevpkaeleklitmkpgkwyerqgm	301		
QY	308 EQTTNNISTKFGDDGYVYAOIRPVTRINDESRVVDVEYYIDPVHPVYVRRINFTGNPKTQ	367		
Db	302 tavlgeiqnrmgsagayseisvqplpnaetktvdfvlhiepggrkiyvneihitgnkntr	361		
QY	368 DEVLREMRLEGALASNOIKQLSARLMRTGFFKHVTVDRPVNSPDQVDVNFVVEEQ	427		
Db	362 devvrelrmeapdytsklgrskervellgyfdnvqfdavplagtdpkvdlmnlalter	421		
QY	428 PSGSSTIAAGYSQSGVTFQFQVQNNFMCTGKHVNASFSESTREVSILGTMNPNFTVN	487		
Db	422 stgslslsagwvqdtglvmsagvsqdnlfqgksaalrsrsktllngslsfddpyftad	481		
QY	488 GVSQSLSGYR--KTKYDNKNISNVYDLSYSGSLSYGYPIDENQRTSFGNLNADNTKLHGG	545		
Db	482 gvslygdyvvgkafprkastsikqykttagagirmsvpvteydrvnfglvaehltvn--	539		
QY	546 RFMGISNVKQLMADGGKIQVDNNGIPDFKHDTTYNAILGNWYSSLDLPVFTQGNHSHV	605		
Db	540 ---tynkaphyadffikkygktdg--tdgsfkglwlykgtvgwgrnktksalwptgrlytg	595		
QY	606 DLTVGF-GDK-THQKVYQGNIRYRPFIKKSVLR-----GYAKLGYG--NNLPFYEYFAG	656		
Db	596 naeialpgsklqysathnqtfflpskftlmlggevgiag-gygrtkeipffenfygg	654		
QY	657 GYGSVRGYDOSSLGPRSQAYLTARRGQOTFLGEV--GGNALATFGESELPLPFKGDWI	714		
Db	655 glgsvrgyesgtlgpk-----vydeygekisyggnkkanvsaeellfmpgskad-a	703		
QY	715 DOVRPVIETEGGVFTTCMDKQOTIDLTFQKDPQATAEQNAKANRPLLTQDKQLRYSAG	774		
Db	704 rtrvrlsfadagsvwdgkytyddnssat-----ggvrqniygagntkstkftneilrysag	758		
QY	775 VGATWPTPIGLSISYAKPLNKKQNDQDTDTVQFGISVF	813		
Db	759 gavlwlspgpmkfsayayplkkkpedeqirfqfqlgttf	797		
RESULT	8			
ID	AAB23784			
XX	AAB23784	standard; Protein; 797 AA.		
AC	AAB23784;			
XX				
DT	12-JAN-2001	(first entry)		
XX				
DE	Neisseria meningitidis serogroup B amino acid sequence.			
XX				
CG motif; complete Freund's adjuvant; phosphorothioate; immunogenic; Neisseria antigen; Neisseria meningitidis; Neisseria gonorrhoeae; bactericidal; antibacterial; vaccine; immunostimulatory; infection; immune response.				
Neisseria meningitidis.				
WO200050075-A2.				
31-AUG-2000.				
09-FEB-2000; 2000WO-IB00176.				
26-FEB-1999; 99US-0121792.				
(CHIR-) CHIRON SPA.				
Grandi G, Rappuoli R, Giuliani M, Pizza M;				
WPI; 2001-015529/02.				
Immunogenic composition useful for stimulating an immune response in a mammal against Neisseria infection, comprises Neisseria antigen and an adjuvant composition comprising an oligonucleotide with a CG motif -				
Claim 22; Page 32; 39pp; English.				
The present invention describes an immunogenic composition (I) comprising a Neisseria antigen and an adjuvant composition comprising an oligonucleotide comprising at least 1 CG motif. Also described is an adjuvant composition (II) comprising an oligonucleotide which comprises at least 1 CG motif and a complete Freund's adjuvant (CFA), where the oligonucleotide preferably comprises at least one phosphorothioate bond. AAA92359 to AAA92385 represent specifically claimed oligonucleotides of the present invention. (I) is useful for stimulating an immune response in a mammal, preferably a human, against Neisseria infection, preferably Neisseria meningitidis infection and in the manufacture of a medicament for inducing a protective immune response in a mammal. The present sequence represents the claimed Neisseria meningitidis serogroup B amino acid sequence disclosed in GB-9928197.4, which is given in the present invention.				
Sequence 797 AA;				
Query Match	26.5%;	Score 1115;	DB 22;	Length 797;
Best Local Similarity	32.5%;	Pred. No. 8.8e-67;		
Matches 266;	Conservative 154;	Mismatches 361;	Indels 38;	Gaps 14;
QY	10 QVSAMTMAVMVMNSTHAQAADFMANDITITGLQRTVIESLQSVLPFRLGQVSVENQLADG	69		
Db	2 ktkqiasalmmlgisplaladftiqdirvegqrtepstvfnylpkvvgdyndthgsai	61		
QY	70 VKALYATGNFSDVOVYHQEGRIIYQVTERPLIAEINFEGRNRLIPKEGLQGLKNAGLAVG	129		
Db	62 ikslyatgfdvrvetadgqllltvterptgslntgskmlqndaknklesfglags	121		
QY	130 QPLKQATVQMIETELTNOYISOGYNTETVKTMLDGNRVKLDMTFAEGKPARVVDINI	189		
Db	122 qyfnqatlnqavaglkceylgrgklniqitpkvtklarnrvdiditidegksakitdef	181		
QY	190 IGNQHFSADLIDVLAIKDKNI-NPLSKADRYTQEKLVTSLENLRKAKYLVFRFFIKD	248		
Db	182 egngvysdrklmrqmsltggtwtlrsnqfneqfagdmekvtfdygnngyfdfrild	241		
QY	249 AKLNINEDKNRIFVEISLHGEQYRFGQOTFLGNLT-YTQAELEALLKFAEGFSQAML	307		
Db	242 tdiqtnektktikitvhggrfwgksiegdtnevpkaeleklitmkpgkwyerqgm	301		
QY	308 EQTTNNISTKFGDDGYVYAOIRPVTRINDESRVVDVEYYIDPVHPVYVRRINFTGNPKTQ	367		
Db	302 tavlgeiqnrmgsagayseisvqplpnaetktvdfvlhiepggrkiyvneihitgnkntr	361		
QY	368 DEVLREMRLEGALASNOIKQLSARLMRTGFFKHVTVDRPVNSPDQVDVNFVVEEQ	427		
Db	362 devvrelrmeapdytsklgrskervellgyfdnvqfdavplagtdpkvdlmnlalter	421		
QY	428 PSGSSTIAAGYSQSGVTFQFQVQNNFMCTGKHVNASFSESTREVSILGTMNPNFTVN	487		
Db	422 stgslslsagwvqdtglvmsagvsqdnlfqgksaalrsrsktllngslsfddpyftad	481		
QY	488 GVSQSLSGYR--KTKYDNKNISNVYDLSYSGSLSYGYPIDENQRTSFGNLNADNTKLHGG	545		
Db	482 gvslygdyvvgkafprkastsikqykttagagirmsvpvteydrvnfglvaehltvn--	539		
QY	546 RFMGISNVKQLMADGGKIQVDNNGIPDFKHDTTYNAILGNWYSSLDLPVFTQGNHSHV	605		
Db	540 ---tynkaphyadffikkygktdg--tdgsfkglwlykgtvgwgrnktksalwptgrlytg	595		
QY	606 DLTVGF-GDK-THQKVYQGNIRYRPFIKKSVLR-----GYAKLGYG--NNLPFYEYFAG	656		
Db	596 naeialpgsklqysathnqtfflpskftlmlggevgiag-gygrtkeipffenfygg	654		
QY	657 GYGSVRGYDOSSLGPRSQAYLTARRGQOTFLGEV--GGNALATFGESELPLPFKGDWI	714		
Db	655 glgsvrgyesgtlgpk-----vydeygekisyggnkkanvsaeellfmpgskad-a	703		
QY	715 DOVRPVIETEGGVFTTCMDKQOTIDLTFQKDPQATAEQNAKANRPLLTQDKQLRYSAG	774		
Db	704 rtrvrlsfadagsvwdgkytyddnssat-----ggvrqniygagntkstkftneilrysag	758		
QY	775 VGATWPTPIGLSISYAKPLNKKQNDQDTDTVQFGISVF	813		
Db	759 gavlwlspgpmkfsayayplkkkpedeqirfqfqlgttf	797		

QY 368 DEVLRRMRQLEGALASNQIKQLSRARLMRTGFFKHVTVTRPVPNSPDQDVNFVBEQ 427  
 Db 362 devrrlrmqmesapydtsklqrskervellgyfdnvqfdavplagtpdkvdlmslter 421  
 QY 428 PSGSSTIAAGYSQGGVTFQFDVSONFMGTGKHVNASFSESRSEREVYSLGTMNPYFTVN 487  
 Db 422 stgslldisagwvqdtglvmsagvsqdnlfgtksaalrskrsktllngslsftdpyftad 481  
 QY 488 GVSQSLSGYYR--KTKYDNKNISVYLDYSGSLSYGYPIDENQIRISFGLNADNTKLHGG 545  
 Db 482 gvslygydygkafprkastsikyktttagagirmsvpvteydrvnfvlvaehltnv-- 539  
 QY 546 RFMGISNVKQLMADGGKIQVDNNGIPDPFKHDYTYNAILGNWYSSLDPRVPTQGMHSV 605  
 Db 540 ---tynkaphyadifkkygktgd-tdgsfkgwlygtvgwgrnktdsalwptrgyitgv 595  
 QY 606 DLTVGF-GDK--THQKVYVQGNIRYRPFYIKKSVLR-----GYAKLGYG--NNLPFYENFVAG 656  
 Db 596 naeialpgsklqyysathnqtwtffplsktftlmjlggevagiag-gygrtkelpfenfygg 654  
 QY 657 GYGSVRGYDQSSSLGPRSQAYLTARRGQOTTGLGEVY--GGNALATFGSELILPLPKGDWI 714  
 Db 655 glgsvrgyesgtlqpk-----vydeygekiysygnkkanvsaeallfmpgpkad-a 703  
 QY 715 DOVRPVFIEGGQVDFDTGMDKQITDLTQFKDPQATAEONAKAANRPLLTDQKOLRYSG 774  
 Db 704 rtvlslfadagsvwdgktydndssat-----ggrvgni ygagnthkstkftnelrysg 758  
 QY 775 VGATWYPIGPLSISYAKPLNKKQNDQTDVTOFOIGSVF 813  
 Db 759 gavtswlspgmkfsyaplkpkpedelqrfqfqlgttf 797

RESULT 9  
 AAU03959  
 ID AAU03959 standard; Protein; 797 AA.  
 XX AAU03959;  
 XX 23-OCT-2001 (first entry)  
 DE Neisseria gonorrhoeae antigenic protein.  
 XX Antigenic protein; pharynx; meningitis; septicaemia; mammalian cell;  
 KW bacterial infection; baculovirus; yeast; Neisseria meningitidis.  
 OS Neisseria gonorrhoeae.  
 XX Key Location/Qualifiers  
 FT Peptide 1..21  
 FT Protein /note= "Signal peptide"  
 FT 22..797  
 FT /note= "Mature N. gonorrhoeae antigen"  
 XX W0200138350-A2.  
 XX 31-MAY-2001.  
 XX 28-NOV-2000; 2000WO-IB01851.  
 XX 29-NOV-1999; 99GB-0028197.  
 PR 09-MAR-2000; 2000GB-0005698.  
 XX (CHIR-) CHIRON SPA.  
 PA (STAT-) STATENS INST FOLKEHELSE.  
 PA Giuliani MM, Piza M, Rappuoli R, Holst J;  
 XX WPI; 2001-381289/40.  
 DR N-PSDB; AAS07279.  
 XX Novel 85 kDa antigen from Neisseria meningitidis and Neisseria  
 PT gonorrhoeae, useful in the manufacture of a medicament for treating and

preventing Neisserial bacteria infection -  
 Claim 1; Page 37-39; 92pp; English.  
 CC The sequence represents a Neisseria gonorrhoeae 85 kDa antigenic protein.  
 CC N. gonorrhoeae is closely related to N. meningitidis, which colonises the  
 CC pharynx, causing meningitis and, occasionally, septicaemia in the absence  
 CC of meningitis. This antigenic protein is useful in the manufacture of a  
 CC medicament for treating or preventing infection due to Neisseria  
 CC bacteria, such as meningitis and septicaemia. It is also useful as a  
 CC diagnostic reagent for detecting the presence of Neisseria bacteria or  
 CC antibodies raised against Neisseria, and as a reagent for raising the  
 CC variety of different expression systems, for example, mammalian cells,  
 CC baculoviruses, plants, bacteria and yeast.  
 XX Sequence 797 AA;  
 SQ

Query Match 26.4%; Score 1111; DB 22; Length 797;  
 Best Local Similarity 32.4%; Pred. No. 1.6e-66;  
 Matches 265; Conservative 155; Mismatches 361; Indels 38; Gaps 14;

QY 10 QVSAMTMAVMVMSTHQAQAFMANDIITGLQRTVIESLSQSLVFLRQGVVSENLADG 69  
 Db 2 kklqasalmvlgisplaladftqdirvegqlrtepstvfnylpkvgydntdthgsai 61  
 QY 70 VKALYATGNFSDVOVYHOEGRIIYQVTERPLIAETNFGNRLIPKEQLEGKLNAGLAVG 129  
 Db 62 kkslyatgffdvrvetadgqlitlvierptgslngakmlqndaikklesfglaqs 121  
 QY 130 QPLQATVQMIETELTNOYISQYYNTEITVQTMLDGNRVKMDMTFAEGKPARVVDINI 189  
 Db 122 qyfnqatlnqavaglkceylgrgklniqtpkvtklarnrvdiditidegksakitdief 181  
 QY 190 IGNOHFSADLIDVLAIKDKNI-NPLSKADRYTQKLVTSLENLRAKYLNAGFVFEIKD 248  
 Db 182 egngvysdrklmrqmsitdeggiwtlrsnqfneqfagdmekvtdfygnngyfdfrild 241  
 QY 249 AKLNEDKNRIFVEISLHEGEYRFGOTQFLGNLT-YTQAELEALLFKAEEGSQAML 307  
 Db 242 tdiqtnedkktqtkitvhggrfrwkvsgiedtnevpkaeleklmlkmpgkyvyrqgm 301  
 QY 308 EQTTNNISTKFGDGYAAQIRPVTRINDESRTVDVEYIIDPVHPVYVRRINFTGNFKTQ 367  
 Db 302 taviqeiqrngmsagayseisvqplpnaetktdfvlhiepgkrkiyvneihitgunktr 361  
 QY 368 DEVLRRMRQLEGALASNQIKQLSRARLMRTGFFKHVTVTRPVPNSPDQDVNFVBEQ 427  
 Db 362 devrrlrmqmesapydtsklqrskervellgyfdnvqfdavplagtpdkvdlmslter 421  
 QY 428 PSGSSTIAAGYSQGGVTFQFDVSONFMGTGKHVNASFSESRSEREVYSLGTMNPYFTVN 487  
 Db 422 stgslldisagwvqdtglvmsagvsqdnlfgtksaalrskrsktllngslsftdpyftad 481  
 QY 488 GVSQSLSGYYR--KTKYDNKNISVYLDYSGSLSYGYPIDENQIRISFGLNADNTKLHGG 545  
 Db 482 gvslygydygkafprkastsikyktttagagirmsvpvteydrvnfvlvaehltnv-- 539  
 QY 546 RFMGISNVKQLMADGGKIQVDNNGIPDPFKHDYTYNAILGNWYSSLDPRVPTQGMHSV 605  
 Db 540 ---tynkaphyadifkkygktgd-tdgsfkgwlygtvgwgrnktdsalwptrgyitgv 595  
 QY 606 DLTVGF-GDK--THQKVYVQGNIRYRPFYIKKSVLR-----GYAKLGYG--NNLPFYENFVAG 656  
 Db 596 naeialpgsklqyysathnqtwtffplsktftlmjlggevagiag-gygrtkelpfenfygg 654  
 QY 657 GYGSVRGYDQSSSLGPRSQAYLTARRGQOTTGLGEVY--GGNALATFGSELILPLPKGDWI 714  
 Db 655 glgsvrgyesgtlqpk-----vydeygekiysygnkkanvsaeallfmpgpkad-a 703  
 QY 715 DOVRPVFIEGGQVDFDTGMDKQITDLTQFKDPQATAEONAKAANRPLLTDQKOLRYSG 774

704	rtrvrlsfadagavdgkytddnssst-----ggvrqvnlygagnthkstftnelrysag
775	VGATWTPIGPLSISYAKPLNKKQNDQTDTVQFQIGSVF
759	gavtlwslpgpmkfsyayplkkpkpedeigrfqqlgttf
 RESULT 10 AAU04451 ID AAU04451 standard; Protein; 797 AA. XX AC AC AC AC AC AAU04451; XX XX DT 23-OCT-2001 (first entry) XX XX DE Neisseria meningitidis serogroup A antigenic protein #2. KW Serogroup B antigen; pharynx; meningitis; septicaemia; mammalian cell KW bacterial infection; baculovirus; yeast. XX XX OS Neisseria meningitidis. XX XX FH Key Location/Qualifiers FT Peptide 1..21 FT FT /note= "Signal peptide" FT Protein 22..797 FT FT /note= "Mature N. meningitidis serogroup A antigen" XX XX PN W0200138350-A2. PD 31-MAY-2001. XX XX PF 28-NOV-2000; 2000WO-TB01851. XX XX PR 29-NOV-1999; 99GB-0028197. PR 09-MAR-2000; 2000GB-0005698. PA (CHIR-) CHIRON SPA. PA (STAT-) STATENS INST FOLKEHELSE. XX XX GI Giuliani MM, Pizza M, Rappuoli R, Holst J; PI WPI; 2001-381289/40. DR XX XX XX Novel 85 kDa antigen from Neisseria meningitidis and Neisseria gonorrhoeae, useful in the manufacture of a medicament for treating and preventing Neisserial bacteria infection - PT PT PT Claim 1; Page 39-40; 92pp; English. PS XX The sequence represents a Neisseria meningitidis serogroup A 85 kDa antigenic protein. Neisseria meningitidis colonises the pharynx, causing meningitis and, occasionally, septicaemia in the absence of meningitis. This antigenic protein is useful in the manufacture of a medicament for treating or preventing infection due to Neisseria bacteria, such as meningitis and septicaemia. It is also useful as a diagnostic reagent detecting the presence of Neisseria bacteria or antibodies raised against Neisseria, and as a reagent for raising the antibodies. The Neisseria nucleotide sequences can be expressed in a variety of different expression systems, for example, mammalian cells, baculoviruses, plant bacteria and yeast. CC Note: There are two versions of this sequence displayed in the specification (see AAU03958). XX Sequence 797 AA; SQ	

Query Match 26.4%; Score 1111; DB 22; Length 797;  
Best Local Similarity 32.4%; Pred. No. 1.6e-66;  
Matches 265; Conservative 155; Mismatches 361; Indels 38; Gaps 14;

OY 10 QVSAMTNVMVMSSTHAQAADFMANDTITGLQRVTIESQSLVPFRLGVQVSENQLADG 69  
:: : |:| :: : |||| |::| :| :| :| :| :

Db	2	klkqiasalmvlgisplaladaftiqdirveglqrtepstvfnylpvkvqgdyndthgsai	61
Qy	70	VKALYATCNFSDVOVYHQEGRIIQVTERPLAIEINPEGNRLIPKEGLQSLKAGLAVG	129
Db	62	iksiyatgfddvrvetadgqllltvtertigslnltgkmlqndaikknllesfqlaqs	121
Qy	130	OPLKQATVQMIELTETLNOYISQGYNTEITVKOTMLDGNRVKLDMTFAEGKPARVVDINI	189
Db	122	qyfnqatlngavaglkkeeylgrgklnlqipkvtkklarnrvdiditidegksakitdief	181
Qy	190	IGNQHFSDADLIDVLAIKDNKI-NPLSKADRYTQEKLVTSLENLKAKYLNAGVFRFEIKD	248
Db	182	egnvyssorklmrgmslttegiwtclwlrnsqfneqkfaqdmekvtdfyqngvdfdrild	241
Qy	249	AKLINEDKNRIFVEISLHGEQVRFOTGFLGNLT-VYQAELEALLKFAEGLFSGQAML	307
Db	242	tdigtneedkktikittvheggrfrfwkvsiegdtnevpkaeleklmtkpgkwyerqgm	301
Qy	308	EQTNNISTKPGDDGYVYAAQIRPVTRINDESRTVDVEYYIDPVHPVYVRRINFTGNFKTQ	367
Db	302	tavigeiqnrmgsagysayseisvqplnaetektvdfvlhiepgrkiyvneihitgnkntr	361
Qy	368	DEVLRERMRQLEGALASNOKIQLSRARLMRTGTFPKKHVTVDRPVNSPDQVDVNFVVEEQ	427
Db	362	devvrilrlimesapydtsklrskervellgyfdnvqfdavplagtpdkvldnmlter	421
Qy	428	PSGSTTAAGYSQGGVTFQFDVSONFMGTGKHVNASFRRSSTREYVSLGTMNPVFTVN	487
Db	422	stgldlisdagvwqdtgvlmsagvsqdnlfgtgksaalrasrktktlngslsfcdpyftad	481
Qy	488	GVSQSLSGYYR-KTKYDNKNISNYVLDYSGSLSYGYPIDENQIRISFGLNADNTKLHGG	545
Db	482	gvslygydvgykafprkastsikqykttagagirmsvpvteydrvnfglvaehltvn--	539
Qy	546	RFMGISNVKQLMADGGKIQVDNNGIPDPFKHDYTYNAILGNVYSSLDPRVPFPGQSHSV	605
Db	540	---tynkaphyadikkyktdtg-tdgsfkwylygtvgwgnktdsalwprgyltg	595
Qy	606	DLTWGF-GDK-THQKVVYQGNIVRPFTKKSVLR-----GVAKLYG--NNLPVENFYAG	656
Db	596	naelialpgsklqysathnqtwffplskttflmlggevlag-gygrtkelpiffenfygg	654
Qy	657	GYGSVRGYDOSSGLPRSOAVLTARRGQOTTLGVEV--GGNALATFGSELILPLPFKGDWI	714
Db	655	glgsrvrgesgtlqpk-----vydeygekisyggnkkanvsaellfmpgakk-a	703
Qy	715	DQVRPVFIIEGGQVFDTHGMDKQTDILTQFKDFOQTAEQNAKAAANRPDLLTQDKQLRYSAG	774
Db	704	rtvlslsfadagsvwdktyddnsssat-----ggrvqniygagnthkstfnelrysag	758
Qy	775	VGATWYTPIGPLTSYSAKPLKNKONDQTDVQFOIGSVF	813
Db	759	gavtwslpplgmksfayplkkpdeieiqrfqfqlgttf	797
RESULT	11		
AAAB84746			
ID	AAAB84746	standard; Protein; 797 AA.	
XX	AAAB84746;		
AC	AAAB84746;		
XX			
DT		17-SEP-2001 (first entry)	
XX			
DE		Amino acid sequence of a Neisseria serogroup A protein.	
XX			
KW		Serogroup A protein; outer membrane protein; Neisserial infection;	
KW		vaccine.	
XX			
OS		Neisseria meningitidis.	
XX			
FH		Location/Qualifiers	
FT	Peptide	1..21	
FT		/note= "signal peptide"	













GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 09:54:36 ; Search time 37.39 seconds  
(without alignments)  
531.105 Million cell updates/sec

Title: US-09-701-711-4  
Perfect score: 4202  
Sequence: 1 MRNSYFKGFQVSAMTAVMM.....LNKKQNDQDTDTVOFQIGSVF 813

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1052.5	25.0	797	3	US-08-433-522A-2
2	1052.5	25.0	797	3	US-08-433-522A-4
3	1052.5	25.0	797	3	US-08-433-522A-6
4	1052.5	25.0	797	3	US-09-135-166-2
5	1052.5	25.0	797	3	US-09-135-166-4
6	1052.5	25.0	797	3	US-09-135-166-6
7	1052.5	25.0	797	3	US-08-942-046-2
8	1052.5	25.0	797	4	US-08-942-046-4
9	1052.5	25.0	797	4	US-08-942-046-6
10	1047.5	24.9	793	3	US-08-433-522A-10
11	1047.5	24.9	793	3	US-09-135-166-10
12	1047.5	24.9	793	4	US-08-942-046-10
13	1042	24.8	792	3	US-08-433-522A-8
14	1042	24.8	792	3	US-09-135-166-8
15	1042	24.8	792	4	US-08-942-046-8
16	146.5	3.5	2123	4	US-08-968-685A-10
17	142	3.4	2199	5	PCT-US95-11684-2
18	141	3.4	1599	2	US-08-617-697-9
19	138	3.3	1536	1	US-08-038-682-2
20	138	3.3	1536	1	US-08-302-832-2
21	138	3.3	1536	2	US-08-530-198-2
22	138	3.3	1536	2	US-08-469-880-2
23	138	3.3	1536	2	US-08-728-470-2
24	138	3.3	1536	2	US-08-617-697-2
25	138	3.3	1536	4	US-08-719-641-2
26	136.5	3.2	2314	4	US-09-268-347-49
27	136	3.2	990	4	US-09-627-376-7

28	134.5	3.2	1178	1	US-08-446-486-5	Sequence 5, Appli
29	134.5	3.2	1178	1	US-08-463-308-5	Sequence 5, Appli
30	134.5	3.2	1182	1	US-08-349-867-34	Sequence 34, Appl
31	134.5	3.2	1182	2	US-08-598-305A-34	Sequence 34, Appl
32	134.5	3.2	1188	1	US-08-239-476-34	Sequence 34, Appl
33	134.5	3.2	1188	2	US-08-639-923A-34	Sequence 34, Appl
34	134.5	3.2	1188	5	PCT-US95-05431-34	Sequence 34, Appl
35	134	3.2	2048	4	US-09-268-347-48	Sequence 48, Appl
36	132	3.1	905	4	US-09-074-658-70	Sequence 70, Appl
37	130	3.1	969	1	US-07-671-817A-4	Sequence 4, Appli
38	130	3.1	1177	1	US-07-828-788A-8	Sequence 8, Appli
39	130	3.1	1177	1	US-07-920-085-2	Sequence 2, Appli
40	130	3.1	1177	5	PCT-US92-11337-8	Sequence 8, Appli
41	130	3.1	1177	6	5169629-2	Patent No. 5169629
42	129.5	3.1	2089	1	US-08-418-893D-23	Sequence 23, Appl
43	129.5	3.1	2089	1	US-08-418-893D-24	Sequence 24, Appl
44	128.5	3.1	682	3	US-08-613-009A-10	Sequence 10, Appl
45	128.5	3.1	702	2	US-08-867-941-25	Sequence 25, Appl

ALIGNMENTS

RESULT 1  
US-08-433-522A-2  
; Sequence 2, Application US/08433522A  
; Patent No. 6013514  
; GENERAL INFORMATION:  
; APPLICANT: CHONG, Pele  
; APPLICANT: THOMAS, Wayne  
; APPLICANT: YANG, Yan Ping  
; APPLICANT: LOOMORE, Sheena  
; APPLICANT: STA, Dwo Yuan Charles  
; APPLICANT: KLEIN, Michel  
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN  
; NUMBER OF SEQUENCES: 55  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6TH Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/433,522A  
; FILING DATE: 12-SEP-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: STEWART, Michael I  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-434 MIS:jib  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 797 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-433-522A-2

Query Match 25.0%; Score 1052.5; DB 3; Length 797;  
Best Local Similarity 31.6%; Pred. No. 4.5e-76;  
Matches 260; Conservative 160; Mismatches 333; Indels 69; Gaps 20;  
QY 23 STHAQAADFMAADITITGLQRTVIESLQSVLPRLQGVVSENQLADGVKALYATGNFSDV 82

```
Db 14 TTTFAAPFAVKDIKVDGQGLDEQIIRASLPVRAGQRTDNDVANIYRSLFVSGRFDV 73
Qy 83 QVYHOEGRI-IYQVTERPLIAEINPEGNRLPKEGLOGLKNAGLAVGQPLKQATVQOMIE 141
Db 74 KA-HOEGDVLVSVVAKSIISDVKIKGNSVPTPEALKONLDANGFKVGDVLIKELNEFA 132
Qy 142 TELNNOYISQYNYTEITVTKQMLDGNRVKLDMTFAEGKPARVVDINIGNQHFSDADLI 201
Db 133 KSVKEHYASVGRYNATPEIVNTLPNNRAELIIOINEDDKAKLASLTFKGNESVSSLTQ 192
Qy 202 DVLAIKONKINPLSKADRYTQELVTSLENIRAKYLNAGFVRFEIKDAKLINEDKNRIF 261
Db 193 EOMELQPSDWKWL-WGNKFEQAQFEKDLQSIYRDYLLNNGYAKAQITKTDVQNDKTKVN 251
Qy 262 VEISLHEGEQYRFQGTQFLNLTQAELEALLK-FKAEEGFSOAMLEPQTNNISTKFGD 320
Db 252 VTIDVNEGLQYDLRSARIIGNLGMSAELEPLLSALHLNDFRSDIADVENAKAKLGE 311
Qy 321 DGYAAQIRPVTRINDESRTVDVEYIDPVHPVYVRRINFTGNFKTQDEVLRREMRLQEG 380
Db 312 RGYSATVNSVPDFEDDANKTLAITLVVDAGRRLTVRQLRFEQNTVSADSTLRQEMRQEG 371
Qy 381 ALASNOKIOLSRALMRGFEKHHVTVDTR--PVPNSPDQVDVNFVEEOPSGSSTIAAGY 438
Db 372 TWYNSQLVELGKIRLDRGTFFE--TVENRIDPINGSDEVDVYKVKERNGTGINSFGIGY 429
Qy 439 SQSGGVTFQFDVSONNFMGTGKHYNASFSRSETREVYSLGMTNPFYTVNGVQSLSGYR 498
Db 430 GTESSISYQASVKODNFGTGAAYSIAGTKNDYGTSVNLGYTEPYFKDGV--SLGNGVF 487
Qy 499 KTKYDNK---NISNYLDSYGGISYGYPIDENQRISEGL-----NADNTKLHGGRPMGI 550
Db 488 FENDNKSDDTSSNYKRTYTSNVTGLFPVNNENNSYVGLGHTYNTKISNFALEYNRNLYI 547
Qy 551 SNVKQLMADGGKIQVDNNGIPDFKHDTTYNAIILGWYSSLDPRVPFTQGMHSHVD--LT 608
Db 548 QSMK-----FKNGIKTNDDFS-----FGWYNSLNRGYFPTKGVKASLGGRVT 592
Qy 609 VFGDKTHQKVYQCNIRYRP-----IKKSVLRGYAKLGYGN-NLPFFYENFYAGGYGS 660
Db 593 IPGSDKNVYKLSADVOGFYPLDRHLVWVSAKASAGYAN-GFCNKRLPFFQYTYAGGIGS 651
Qy 661 VRGYDQSSGLPSQAYLTARQOQTTIG-----EVVGNALATFGESELILPLPFKGDW 713
Db 652 LRFPAYGSIQPN---IYAEYNGSGTGTFKKITSSDVIIGNATATASAEELIVPTPFVSDK 708
Qy 714 I-DOVRPVIFTEGGQVDDTT-GMDKQITDLTQFKDPQATAEQNAKANRRPLLTQDKOLRY 771
Db 709 SONTVRSLEFDAASVWNTKWSKXGLESVLLK-----RLPDYKSSKIRA 755
Qy 772 SAGVGATWYTPIGPLSISYAKPLNKKQNDQTDTVQFOIGSVF 813
Db 756 STGVGFQWQSPIGPLVFSYAKPKIKYENDDDVEQFOFSGISGF 797
```

## RESULT 2

```
US-08-433-522A-4
; Sequence 4, Application US/08433522A
; Patent No. 6013514
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
```

```
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433,522A
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-434 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 797 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-433-522A-4
```

```
Query Match 25.0%; Score 1052.5; DB 3; Length 797;
Best Local Similarity 31.6%; Pred. No. 4.5e-76;
Matches 260; Conservative 160; Mismatches 333; Indels 69; Gaps 20;

Qy 23 STHAAADFMANDITITGLQRTVETIESLQSVLPFRIGVSVSENOADGVKALYATGNFSDV 82
Db 14 TTTFAAPFAVKDIKVDGQGLDEQIIRASLPVRAGQRTDNDVANIYRSLFVSGRFDV 73
Qy 83 QVYHOEGRI-IYQVTERPLIAEINPEGNRLPKEGLOGLKNAGLAVGQPLKQATVQOMIE 141
Db 74 KA-HOEGDVLVSVVAKSIISDVKIKGNSVPTPEALKONLDANGFKVGDVLIKELNEFA 132
Qy 142 TELNNOYISQYNYTEITVTKQMLDGNRVKLDMTFAEGKPARVVDINIGNQHFSDADLI 201
Db 133 KSVKEHYASVGRYNATPEIVNTLPNNRAELIIOINEDDKAKLASLTFKGNESVSSLTQ 192
Qy 202 DVLAIKONKINPLSKADRYTQELVTSLENIRAKYLNAGFVRFEIKDAKLINEDKNRIF 261
Db 193 EOMELQPSDWKWL-WGNKFEQAQFEKDLQSIYRDYLLNNGYAKAQITKTDVQNDKTKVN 251
Qy 262 VEISLHEGEQYRFQGTQFLNLTQAELEALLK-FKAEEGFSOAMLEQTTNNISTKFGD 320
Db 252 VTIDVNEGLQYDLRSARIIGNLGMSAELEPLLSALHLNDFRSDIADVENAKAKLGE 311
Qy 321 DGYAAQIRPVTRINDESRTVDVEYIDPVHPVYVRRINFTGNFKTQDEVLRREMRLQEG 380
Db 312 RGYSATVNSVPDFEDDANKTLAITLVVDAGRRLTVRQLRFEQNTVSADSTLRQEMRQEG 371
Qy 381 ALASNOKIOLSRALMRGFEKHHVTVDTR--PVPNSPDQVDVNFVEEOPSGSSTIAAGY 438
Db 372 TWYNSQLVELGKIRLDRGTFFE--TVENRIDPINGSDEVDVYKVKERNGTGINSFGIGY 429
Qy 439 SQSGGVTFQFDVSONNFMGTGKHYNASFSRSETREVYSLGMTNPFYTVNGVQSLSGYR 498
Db 430 GTESSISYQASVKODNFGTGAAYSIAGTKNDYGTSVNLGYTEPYFKDGV--SLGNGVF 487
Qy 499 KTKYDNK---NISNYLDSYGGISYGYPIDENQRISEGL-----NADNTKLHGGRPMGI 550
Db 488 FENDNKSDDTSSNYKRTYTSNVTGLFPVNNENNSYVGLGHTYNTKISNFALEYNRNLYI 547
Qy 551 SNVKQLMADGGKIQVDNNGIPDFKHDTTYNAIILGWYSSLDPRVPFTQGMHSHVD--LT 608
Db 548 QSMK-----FKNGIKTNDDFS-----FGWYNSLNRGYFPTKGVKASLGGRVT 592
Qy 609 VFGDKTHQKVYQCNIRYRP-----IKKSVLRGYAKLGYGN-NLPFFYENFYAGGYGS 660
```

```
Db 593 IPGSDNKYKLSADYQGFPLDRDLHWVVSASAKAGYAN-GFGNKRPLPYQTYTAGGIGS 651
Qy 661 VRGYDOSSLGPRSQAYLTARRQOQTTLG-----EVVGGNALATFGSELILPLPKGDW 713
Db 652 LRGFAYSGIPNA---IYAEYNGSGTGFFKISSDVGIGNALATASAEIIVPTPFVSDK 708
Qy 714 I-DQVRPVIIFEGGVFDIT-GMDKQTDIDLTKQKQATAEQNAKAAANRPLLTQDKQLRY 771
Db 709 SONTVRTSLFVDAASVWNTKWKSDKNGLSDVLK-----RLPDYKSKSRIRA 755
Qy 772 SAGVGATWTPIGPLSISYAKPLNKKQNDQTDITVQFQIGSVF 813
Db 756 STGVGFQWQSPIGPLVFSIAKPKIKKYENDDDVEQFQFISIGGSF 797

RESULT 3
US-08-433-522A-6
; Sequence 6, Application US/08433522A
; Patent No. 6013514
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433,522A
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-434 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 797 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-433-522A-6

Query Match 25.0%; Score 1052.5; DB 3; Length 797;
Best Local Similarity 31.6%; Pred. No. 4.5e-76;
Matches 260; Conservative 160; Mismatches 333; Indels 69; Gaps 20;

Qy 23 STHAQAQAFMNDITITGLQRTVTSLSQVLPFRGQVVSLENOLADGVKALYATGNFSDV 82
Db 14 TTTVFAFPFAKDVIRVDGVDGQDQIRASLPVRAQQRVTDNVANIVRSLEVSGRFDDV 73
Qy 83 QVYHQEGRI-IQVTERPLIAINEFGRNPLIPKEGLEKNAGLAVGQPLQATVQMIE 141
Db 74 KA-HQEGDVLVVSVAKSIISDVKIKGNSVIPTKALKQNLDAKNGFKVGDVLIREKLINEFA 132
```

```
Qy 142 TELTNOYISQYNTETIVKQTMLDGNRVRKLDMTFAEGKPARVVDINIIGNOHFSADLI 201
Db 133 KSVKEHVASVGRYNATVEPVTNPLPNRAEILIQINEDDKAKLASLTFKGNESVSSTLQ 192
Qy 202 DVLAIKDNKINPLSKADRYTOEKLVTSLNLRKAYLNAGEVFRFEIKADKININEDKNRIF 261
Db 193 EQMELQPDQSWKIL-WGNKFEQAQFEDLQSIDRYLNNGYAKAQITKTVDQLNDEKTKVN 251
Qy 262 VEISLHEGQYRFGQTOFLGNLTFTQAELEALK-FKAEEGFSQAMLEQTTNNISLTKFGD 320
Db 252 VIDVNEGLOYDLRSARIIGNLGGMSAELEPLLSALHLNDTFRSDIADVENAIKAKLGE 311
Qy 321 DGYIAQIRPVTRINDESRVDVEYIDPVHPVYVRRINFTGNFKTQDEVLRRMRQLEG 380
Db 312 RGVGSATVNSVPDQDANKTLAITLVVDAGRRLTVRQLRPEGNTVSADSLRQEMRQEG 371
Qy 381 ALASNOKIOLSRARLMRTGFFKHVTVDTTR--PVPNSPDQVDVNVFVEEQSGSSTTAAGY 438
Db 372 TWYNSQLVELGKIRLDRDTGFFE--TVENRIDPINGSNDEVDVYVYKVRNTGSGINEGIGY 429
Qy 439 SQSGGVTFQFDYSONNFMGTGKHVNASFESKRETSRETVYSLGTMTPYFTVNGVSOSLSGYR 498
Db 430 GTESGISYQASVKQDNFLGTGAASVAGTAKNDYGTSVNLGTYTEPYTKDGV--SLGDNVF 487
Qy 499 KTKYDNK---NISNYVLDYSGSLSYGYPIDENQRIISFGL-----NADNTKLHGGRFPMGI 550
Db 488 FENYDNKSDTSSNYKRTTYGSNVTGLGFPVNNENSYVGLGHTYNTKISNFALEVNRLYI 547
Qy 551 SNVKQLMADGGKIQVDNNGIPDKHYDTTYNAILGNWYSSLDPRVPTQGMHSVD--LT 608
Db 548 QSMK-----FKNGIKTNDDEF-----FGWYNSLNRGYPTTKGVKASLAGRVT 592
Qy 609 VGFQDKTHOKVYVQGNIRPF-----TKKSVLRGVAKLGYCN-NLPFYENFYAGSYGS 660
Db 593 IPGSDNKYKLSADYQGFPLDRDLHWVVSASAKAGYAN-GFGNKRPLPYQTTTAGIGS 651
Qy 661 VRGYDOSSLGPRSQAYLTARRQOQTTLG-----EVVGGNALATFGSELILPLPKGDW 713
Db 652 LRGFAYSGIPNA---IYAEYNGSGTGTFKISSDVGIGNALATASAEIIVPTPFVSDK 708
Qy 714 I-DQVRPVIIFEGGVFDIT-GMDKQTDIDLTKQKQATAEQNAKAAANRPLLTQDKQLRY 771
Db 709 SONTVRTSLFVDAASVWNTKWKSDKNGLSDVLK-----RLPDYKSKSRIRA 755
Qy 772 SAGVGATWTPIGPLSISYAKPLNKKQNDQTDITVQFQIGSVF 813
Db 756 STGVGFQWQSPIGPLVFSIAKPKIKKYENDDDVEQFQFISIGGSF 797

RESULT 4
US-09-135-166-2
; Sequence 2, Application US/09135166
; Patent No. 6083743
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/135,166
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 797 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-135-166-2

```

Query Match 25.0%; Score 1052.5; DB 3; Length 797;  
Best Local Similarity 31.6%; Pred. No. 4.5e-76;  
Matches 260; Conservative 160; Mismatches 333; Indels 69; Gaps 20;

QY	23	STHAQADPWANDITITGLQRYTIESLSQVLPFRQGVVSENQIADGVKALYATGNFSDV	82
Db	14	TTTTVAAPFAVKDIRVGVQGDLEQIRASLRVQAGRYTDDVANIIVRSLEVSGRFDV	73
QY	83	QVYHQEGRI--IYOVTERPLIAEINFEGNRLIKPEGLQELGNAGLAVGPOKQATWQMI	141
Db	74	KA-HQEGDVLVSVVAKSIISDVKICGNSVITEALKQNLNDANGFKVGDVLREKLNFEA	132
QY	142	TELTNOYISQYYNTIETVKTQMLDGNRVKLDMTFAEGKPARVVDINIIOHQHFSADLI	201
Db	133	KSYKEHYASVGRYNATVEPVTNLPNNRAELIQUINEDDKAKLASLTFKGENSVSSLTQ	192
QY	202	DVLAIADKNKINPLSKADRYTOEKLVTSLENLRAKYLINAGFVFEIKADKLINEDKNRLF	261
Db	193	EQMELQDPSMWKL-WGNKEFGAOFKDLQSDIRYILNNGYAKAQITKTDVQJLNDSEKTV	251
QY	262	VEISLHEGEQYRGQTFGLNLTYYTQAELEALIK-FKAEEGSQAMLEOTTNNISTKFGD	320
Db	252	VTIDVNEGLQYDLRSARIIGNLGMSAELEPLLSALHLNDTFRSRDIADVENAIRAKLGE	311
QY	321	DGYIIAQIRPVTINDESRVDVEYIYIDPVHPVYVRINFTGNFKTQDDEVLRERMRQLEG	380
Db	312	RGYSATVNSVPDFFDANKTLAHTLVVDAGRRLTVRQLREFGENTVSADSTLRQEMRQEG	371
QY	381	ALASNKIOLSRARMLARTGFFKHVTVDTR--PVPNSPDQVDVNFVVEEOPSSSTIAAGY	438
Db	372	TWYNSQLVELGKIRLDRTEGFE--TVENRIDPINGSNDEVDVYKVKERTGCSINFGICY	429
QY	439	SQSGGVTFQFDYSQNNPMGTGKHVNASFGRSETREYISLGMTNPFYTVNGVVSQSLGYR	498
Db	430	GTESGISYQASVKODNPLGTGAAVSITAGTKNDYGTGVSNTGYEPTFKDGV--SLGGNVF	487
QY	499	KTIYDNK--NISNYVLDSSVGSLSGYPIDENQRISFGL-----NADNTKHLHGGRPMCI	550
Db	488	FENYDNKSQDTSNRYKRTYGSNVTLGFPWNENNSYVGLGHTYKNISFALLEYNRNLIYI	547
QY	551	SNYKQLMADGGKIQVDNNGIPDKHDYTYNATLGNWYSLSDRPFVPTQCMSSHSDV--LT	608
Db	548	QSMK-----FKNGIKTNDFFS-----FGWYNLSNRGYEPTKGVKVASLGRVT	592
QY	609	VFGEDTKHQVYVQGNIIYRPF-----TKKSVLRGYAKLGYGN-NLPFYENFYAGGYGS	660
Db	593	IPGSDNKRYKLSADVGQCFYPLDRHDLHWVYSAKASAGYAN-GFCNKRRLPYQTYTAGGTS	651

```

QY 561 VRGVDQSSLGPRQAYLTARRGQQTTLG-----EWWGNALATGSELILPLPPKGDW 713
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 652 LRGFAYGISGPN-----IYAEGYNGSGTGTFKKISSDVIGGNATATASAEIIVPTFVSDK 708
QY 714 I-DQVRPVIIEGGQVDDTT-GMDKQTIDLTQFKDPQATAEQNAKANRPLLTQDKOLRY 771
   : :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 709 SONTVRTSLFVDAASVWNTKWKSDKNGLSDVLK-----RLPDYKSSRIRA 755
QY 772 SAGYGATWYTPIGPLSISYAKPLNKKQNDQTDVTQVQIGSVF 813
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 756 STGVGFQWQSPIGPLFVSYAKPIKKYENDVDVEQFQSIGGSF 797

RESULT 5
US-09-135-166-4
: Sequence 4, Application US/09135166
: Patent No. 6083743
: GENERAL INFORMATION:
: APPLICANT: CHONG, Pele
: APPLICANT: THOMAS, Wayne
: APPLICANT: YANG, Yan Ping
: APPLICANT: LOOSMORE, Sheena
: APPLICANT: SIA, Dwo Yuan Charles
: APPLICANT: KLEIN, Michel
: TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
: NUMBER OF SEQUENCES: 55
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sim & McBurney
: STREET: 6TH Floor, 330 University Avenue
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5G 1R7
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/135,166
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/433,522
: FILING DATE: 12-SEP-1995
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: STEWART, Michael I
: REGISTRATION NUMBER: 24,973
: REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 595-1155
: TELEFAX: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 797 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-09-135-166-4

```

Query Match 25.0%; Score 1052.5; DB 3; Length 797;  
Best Local Similarity 31.6%; Pred. No. 4.5e-76;  
Matches 260; Conservative 160; Mismatches 333; Indels 69; Gaps 20;

[illegible]









```
Db 74 KA-HOEGDVLVSVVAKSIISDVKIKGNSVIPTTEALKONLDANGFKVGDVILIREKLNFEA 132
Qy 142 TELTNOYISOGYINPEITVKOTMLDGNRVKLDMTFAEKPARYVDINIIGNQHFSDADLI 201
Db 133 KSVKEHYASVGRYNATVEIVNTLPNNRAEILQINEDDKAKLASLTFKGNESVSSSTLQ 192
Qy 202 DVLAIKDNKINPLSKADRYTQEKLVTSLENIIRAKYLNAGVFRFEIKDAKLINEDKNRIF 261
Db 193 BOMELQDPSMWKL-WGNKFEAGFEKDLQSIIRDYILNNGYAKAITKTDVQLNDEKTKVN 251
Qy 262 VEISLHEGEQYRFQTOFLGNLTQAELEALLK-FRAEEGFSQAMLEQTTNNISTKFGD 320
Db 252 VTIDVNEGLQDLRSARIIGNLGMSAELEPLLSALHNDLFRSDIADVENAIKALIGE 311
Qy 321 DGYYYAQIRPVTRINDESRTVDVEYIDPVHPVYVRINFTGNFKTODEVLRREMROLEG 380
Db 312 RGYGSATVNSVPDEFDDANKTLAITLVVDAGRRLTVRLQRFEGNTVSADSTLRQEMRQEG 371
Qy 381 ALASNQIKLSARLMRTGFFKHVTVDTR--PVPNSPDQDVNFVVEQPSGGSSTIAAGY 438
Db 372 TWYNSQLVELGKIRLDRGTFFE--TVENRIDPINGSNDEVDVYKVKERNWTSINFIGY 429
Qy 439 SOSGGVTFOFVDSQNNMGTKHVNASFSETRREYVSLGNTNPFYTVNGVSQSLSGYR 498
Db 430 GTEGSIYSQSVKQDNFLGTGAAYSIAGTKNDYGTSLNGLTGYTEFTYTKDGV--SLGNGVF 487
Qy 499 KTKYDNK---NISNYLDSYGGSLSYGYPIDENQIRISFGL-----NADNTKLHGGRFEMI 550
Db 488 FENYDNKSDTSNNYKRTTYGTSNVTLPFPVNNENSYVGLGHTYKNTSNEALEYNNLYI 547
Qy 551 SNVKQLMADGGKIQVDNNGIPDFKHDTTYNAILGWNYSLLDRPVFTQGNHSHVD--LT 608
Db 548 QSMK-----FKNGIKTNDDFS-----FGWYNSLNRGYFPTKGYKASLGRVT 592
Qy 609 VGFQDKTHQVYVYOGNRYRPP-----IKKSVLRGYAKLGYGN-NLPFYENFYAGGYGS 660
Db 593 IPGSDNKKYKLSADVOGFYPLDRHLVWVSASAGYAN-GFGNKRPLFPFYOTYTAGGIGS 651
Qy 661 VRGYDQSSLPGRSQAYLTARRGOOTTIG-----EVVGNALATFGSELILPLPFKGDW 713
Db 652 LRGPAYGISGPNPNA---IYAEYNGSGTGTFFKIKSSDYGIGNAATASAEELIVPTPFVSDK 708
Qy 714 I-DQVRVIFTEGGQVFDTT-GMDKQITDLTQKDPQATAEQNAKANAANRPLLTQDKOLRY 771
Db 709 SONTVRTSLFVDAASVWNTWKSKNGLESVLLK-----RLPDYKGSRRIRA 755
Qy 772 SAGVGATWYTPIGPLSISYAKPLNKKQNDQTDVTVOFGISVF 813
Db 756 STGVGFQWQSPIGPLVFSYAKPIKKYENDDDVEQFQFSIGGSF 797
```

## RESULT 10

US-08-433-522A-10

; Sequence 10, Application US/08433522A

; Patent No. 6013514

; GENERAL INFORMATION:

; APPLICANT: CHONG, Pele

; APPLICANT: THOMAS, Wayne

; APPLICANT: YANG, Yan Ping

; APPLICANT: LOOSMORE, Sheena

; APPLICANT: SIA, Dwo Yuan Charles

; APPLICANT: KLEIN, Michel

; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN

; NUMBER OF SEQUENCES: 55

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sim &amp; McBurney

; STREET: 6TH Floor, 330 University Avenue

; CITY: Toronto

; STATE: Ontario

; COUNTRY: Canada

; ZIP: M5G 1R7

; COMPUTER READABLE FORM:

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-434 MTS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 793 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-433-522A-10
```

Query Match 24.9%; Score 1047.5; DB 3; Length 793;

Best Local Similarity 31.7%; Pred. No. 1.1e-75;

Matches 259; Conservative 160; Mismatches 334; Indels 65; Gaps 21;

```
Qy 23 STHAAQADFANDITITGLQVTVTIESLQSVLPFRLGQVSENQADGVKALYATGNFSDV 82
Db 14 TTVFAAPPFKDLIRVDGVOGDLEQQIRASLPVRAGORVTDNDVANIVRSLVFSGREDDV 73
Qy 83 QVYHQEGR-IYQVTERPLIAEINFEGRNLRIPKEGLEGLKNAGLVAGQPLKQATVQWIE 141
Db 74 KA-HOEGDVLVSVVAKSIISDVKIKGNSVIPTTEALKONLDANGFKVGDVILIREKLNFEA 132
Qy 142 TELTNOYISOGYINPEITVKOTMLDGNRVKLDMTFAEKPARYVDINIIGNQHFSDADLI 201
Db 133 KSVKEHYASVGRYNATVEIVNTLPNNRAEILQINEDDKAKLASLTFKGNESVSSSTLQ 192
Qy 202 DVLAIKDNKINPLSKADRYTQEKLVTSLENIIRAKYLNAGVFRFEIKDAKLINEDKNRIF 261
Db 193 BOMELQDPSMWKL-WGNKFEAGFEKDLQAIIRDYILNNGYAKAQTITVDQLNDEKTKVN 251
Qy 262 VEISLHEGEQYRFQTOFLGNLTQAELEALLK-FRAEEGFSQAMLEQTTNNISTKFGD 320
Db 252 VTIDVNEGLQDLRSARIIGNLGMSAELEPLLSALHNDLFRSDIADVENAIKALIGE 311
Qy 321 DGYYYAQIRPVTRINDESRTVDVEYIDPVHPVYVRINFTGNFKTODEVLRREMROLEG 380
Db 312 RGYGNTVNSVPDEFDDANKTLAITFVVDAGRRLTVRLQRFEGNTVSADSTLRQEMRQEG 371
Qy 381 ALASNQIKLSARLMRTGFFKHVTVDTR--PVPNSPDQDVNFVVEQPSGGSSTIAAGY 438
Db 372 TWYNSQLVELGKIRLDRGTFFE--TVENRIDPINGSNDEVDVYKVKERNWTSINFIGY 429
Qy 439 SOSGGVTFOFVDSQNNMGTKHVNASFSETRREYVSLGNTNPFYTVNGVSQSLSGYR 498
Db 430 GTEGSIYSQSVKQDNFLGTGAAYSIAGTKNDYGTSLNGLTGYTEFTYTKDGV--SLGNGIF 487
Qy 499 KTKYDNK---NISNYLDSYGGSLSYGYPIDENQIRISFGL-----NADNTKLHGGRFEMI 550
Db 488 FENYDNKSDTSNNYKRTTYGTSNVTLPFPVNNENSYVGLGHTYKNTSNEALEYNNLYI 547
Qy 551 SNVKQLMADGGKIQVDNNGIPDFKHDTTYNAILGWNYSLLDRPVFTQGNHSHVD--LT 608
Db 548 QSMK-----FKNGIKTNDDFS-----FGWYNSLNRGYFPTKGYKASLGRVT 592
Qy 609 VGFQDKTHQVYVYOGNRYRPP-----IKKSVLRGYAKLGYGN-NLPFYENFYAGGYGS 660
Db 593 IPGSDNKKYKLSADVOGFYPLDRHLVWVSASAGYAN-GFGNKRPLFPFYOTYTAGGIGS 651
Qy 661 VRGYDQSSLPGRSQAYLTARRGOOTTIG-----EVVGNALATFGSELILPLPFKGDW 713
Db 652 LRGPAYGISGPNPNA---IYAEYNGSGTGTFFKIKSSDYGIGNAATASAEELIVPTPFVSDK 708
Qy 714 I-DQVRVIFTEGGQVFDTT-GMDKQITDLTQKDPQATAEQNAKANAANRPLLTQDKOLRY 771
Db 709 SONTVRTSLFVDAASVWNTWKSKNGLESVLLK-----RLPDYKGSRRIRA 755
Qy 772 SAGVGATWYTPIGPLSISYAKPLNKKQNDQTDVTVOFGISVF 813
Db 756 STGVGFQWQSPIGPLVFSYAKPIKKYENDDDVEQFQFSIGGSF 797
```

Db 652 LRGFAYGSGPNA-----IYAEHNGCTFNKISSDVIGGNAITTASAEILVPTFPVSDKSQNT 708  
Qy 717 VRPVIFIEGGVDFDT-GMDKOTIDLTOFKDPOATAEQNAKANRPLLTQDKOLRYSGV 775  
Db 709 VRTSLFVDAASVWNTKWSKDNGLSKVLKD-----LPDYGKSSRIIRASTGV 755  
Qy 776 GATWYTPIGPLSISYAKPLNKQNDQDTVQFOIGSVF 813  
Db 756 GFQWSPIGPLVFSYAKPIKKYENDDDVEQFQFSGGSF 793

RESULT 11  
US-09-135-166-10  
; Sequence 10, Application US/09135166  
; Patent No. 6083743  
; GENERAL INFORMATION:  
; APPLICANT: CHONG, Pele  
; APPLICANT: THOMAS, Wayne  
; APPLICANT: YANG, Yan Ping  
; APPLICANT: LOOSMORE, Sheena  
; APPLICANT: SIA, Dwo Yuan Charles  
; APPLICANT: KLEIN, Michel  
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN  
; NUMBER OF SEQUENCES: 55  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6TH Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/135,166  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/433,522  
; FILING DATE: 12-SEP-1995  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: STEWART, Michael I  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 793 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-135-166-10

Query Match 24.9%; Score 1047.5; DB 3; Length 793;  
Best Local Similarity 31.7%; Pred. No. 1.1e-75;  
Matches 259; Conservative 160; Mismatches 334; Indels 65; Gaps 21;

Qy 23 STHQAADFMANDTITGLQRVTIESLQSVLPFRIGVQVSENQJADGWKALYATGNFSDV 82  
Db 14 TTTVFAAPFVFPKDIRVDVGQDLEQQIRASLPVRAGQRTDNDVANIVRSLFVSGREDDV 73  
Qy 83 QVYHQEGRI-IYQTERPLIAEINFEGNRLIPKEGLOGLKAGLAVGQPLKQATVQMI 141  
Db 74 KA-HQEGDVLVVVAAKSIISDVKIGNSVPTALKONLDANGFKVGDVLIREKLNEFA 132

Qy 142 TELTNOYISQYYNTEITVKTMLDGNRVKLDMTFAEGKPARVVDIINIIGNHFSADALI 201  
Db 133 KSVKEHYASVGRYNATVEPIVNTLPNNRAEILIQINEDDKAKLASLTFKGNESVSSSTIQ 192  
Qy 202 DVLAIKDNKINPLSKADRYTQEKLVTSLENRAKYLINAGFVFEIKDAKLINEDKNRIF 261  
Db 193 EQMELQPSDSWML-WGNKFEQAQFEKDLQAIRDYLLNNGYAKAQITKTDOVLNDEKTKVN 251  
Qy 262 VEISLHEGEQYRFQTFGLNLTYYTQAEALLK-FKAEEGFSQAMLEQTTNNISFKFGD 320  
Db 252 VTIDVNEGLOYDLRSARIIGNLGMSAELEPLLSALHLNDTFRSDIADVENAIKALGE 311  
Qy 321 DGYVYQAIRPVTRINDESRTVDVEYIIDPVHPVYVRRINFTGNFKTQDVLRRMRQLEG 380  
Db 312 RGYNTTVNSVPDFDDANKTLAITFVWDAGRLTLRQLRFEGTNSADSTLRLQEMRQOEG 371  
Qy 381 ALASNOKIOLSRARLMRTGFFKHVTVDTR--PVPNSPOVDVNFVVEEQSPSSSTIAAGY 438  
Db 372 TWYNSQLVELGKIRLDRTGFFE--TVENRIDPINGSNDEVVYVYKVKERTGTSINFGIGY 429  
Qy 439 SOSGGVTFQFDVSONNFMTGKGVNASFSRSETREVYSLGTMNPYFTVNGVSQSLSGYR 498  
Db 430 GTESGISYQTSIKQDNFLCTGAASVIAGTKNDYGTSVNLGYTEPYETFKDGV--SLGGNIF 487  
Qy 499 KTKYDNK---NISNYVLDYSGSLSYGYPIDENQRISFGL-----NADNTKLHGGRFMGI 550  
Db 488 FENYDNKSDTSSNYKRTTYGNSVTLGFPVNNNSYVGLGHTYKNKISNFALEYNNRLYI 547  
Qy 551 SNVKQLMADGGKIQVDNNGIPDFKHDTTYNAILGNWYSSLDLRPVFPPTQGMSSHVD--LT 608  
Db 548 QSMK-----FKGNGIKTNDFFS-----FGNNTNSLNRGYFPTKGVKASLGRVT 592  
Qy 609 VGFGDKTHQVV--YQG-----NIYRPFKTSVLGRYAKLGYGN-NLPYENFYAGGYGS 660  
Db 593 IPGSDNKYKLSADVQGFYPLDRHRWVVSASAKAGYAN-GGNKRLPYQIYTAGGIGS 651  
Qy 661 VRGYDOSSLGRPSQAYLTARRQQT---TLGEVVGGNALATFGSELILPLPKGDWI-DQ 716  
Db 652 LRGFAYGSGPNA---IYAEHNGCTFNKISSDVIGGNAITTASAEILVPTFPVSDKSQNT 708  
Qy 717 VRPVIFIEGGVDFDT-GMDKOTIDLTOFKDPOATAEQNAKANRPLLTQDKOLRYSGV 775  
Db 709 VRTSLFVDAASVWNTKWSKDNGLSKVLKD-----LPDYGKSSRIIRASTGV 755  
Qy 776 GATWYTPIGPLSISYAKPLNKQNDQDTVQFOIGSVF 813  
Db 756 GFQWSPIGPLVFSYAKPIKKYENDDDVEQFQFSGGSF 793

RESULT 12  
US-08-942-046-10  
; Sequence 10, Application US/08942046  
; Patent No. 6264954  
; GENERAL INFORMATION:  
; APPLICANT: CHONG, Pele  
; APPLICANT: THOMAS, Wayne  
; APPLICANT: YANG, Yan Ping  
; APPLICANT: LOOSMORE, Sheena  
; APPLICANT: SIA, Dwo Yuan Charles  
; APPLICANT: KLEIN, Michel  
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN  
; NUMBER OF SEQUENCES: 55  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6TH Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

; SOFTWARE: PatentIn Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
;
; APPLICATION NUMBER: US/08/942,046
;
; FILING DATE:
;
; CLASSIFICATION: 435
;
; PRIOR APPLICATION DATA:
;
; APPLICATION NUMBER: 08/433,522
;
; FILING DATE: 12-SEP-1995
;
; CLASSIFICATION: 435
;
; ATTORNEY/AGENT INFORMATION:
;
; NAME: STEWART, Michael I
;
; REGISTRATION NUMBER: 24,973
;
; REFERENCE/DOCKET NUMBER: 1038-732 MIS:jb
;
; TELECOMMUNICATION INFORMATION:
;
; TELEPHONE: (416) 595-1155
;
; TELEFAX: (416) 595-1163
;
; INFORMATION FOR SEQ ID NO: 10:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 793 amino acids
;
; TYPE: amino acid
;
; TOPOLOGY: linear
;
; MOLECULE TYPE: protein
;
; DS-08-942-046-10

```

Query Match	24.98;	Score 1047.5;	DB 4;	Length 793;
Best Local Similarity	31.7%;	Pred. No. 1.le-75;		
Matches	259;	Conservative 160;	Mismatches 334;	Indels 65; Gaps 21;
QY	23	STHAQAADFMANDJITITGLORVTTESLSQSVLPFRLLGOVSVENQOLADGVKALYATGNPSDV	82	
Db	14	TTTVFAAPFPKDIRVGVQGDLEQQIRASLVPVRAGORVTDNDVANIYRSFLVSGREDDV	73	
QY	83	QVYHGEGR1-IYQVTERPLIAEINFEGRNLPIKEGLEKNAGLVCQPLKQATVOMIE	141	
Db	74	KA-HQEGDVLVSVVAKSIISDVKIKGNSVPTALQNLNDANGFKVGDVLIKEKLENEFA	132	
QY	142	TELTNQVISQGYNTETTVKQTMLDGNRVKLDMTFAEGKPARVVDINILIGNOHSADALI	201	
Db	133	KSVKEHYASVGRYNAVTEPIVNTLPPNRAEILIOINEDDRAKASLTFKGNSEVSSSTLQ	192	
QY	202	DVLAIDKNKINPLSKADRYTOEKLVITSLLENRAKYLNAGFVFEIKDAKLININEDKNRIF	261	
Db	193	EOHELQPDSSWKLL-WGNKFGAOPKDLQAIRDYILNNGYAKAQTITKTDVOLNDEKTKVN	251	
QY	262	VELSLHEGEQYRFQQTQFLGNLTVTOALEALKL-FRAEEGFSOAMLEQITNNISTKFGD	320	
Db	252	VTIDVNEGLOYDURSARIILGNLSAELEPLISALHLNDFTRKSDIADVENAIKALGE	311	
QY	321	DGYVYAOIRPVTRINDESRVDVEYYIDPVHPVYVRRINFTGNFKQTODEVLRREMRQLEG	380	
Db	312	RGYGNTVNSVPDFDDANKTLAITFVVDAGRRLTVRQLRFEQNTVSADSTLRQEMRQOEG	371	
QY	381	ALASNOKIOLSRALMETGFEKHKVTVDTDR--PVPNSPDQVDVNVFVEQPSGCSSTIAGY	438	
Db	372	TWTSNSQLVEGKIRLDRGTGFE--TVENRIDPINGSNDEVDVYVKKERNTGCSINFGIGY	429	
QY	439	SQSGGVTFQPDVSONKFMGTCGHVNASFSRSETREVYSLGMTNPFYFVNGVSQSLSGYYR	498	
Db	430	GTESGISYQFSIKQDNFLGTGAASVIAGTKNDYCTSVNLGYTEPYFFTKDGV--SLGNGNIF	487	
QY	499	KTYDKN--NISNYVLDSYGGSLSYGYPTDENQRIISFGL-----NADNTKLGHGRRPMGI	550	
Db	488	FENYDNSKSDTSSNYKRTYTGSVNTLGFPPYNNNSYVYGLGHTYKNKISNFALEYNNRLYI	547	
QY	551	SNYKQLMADGGKIQVDNNGPIPDFKHDYTVTYNAILGNWYSSLDRPVPFTQGMSSHVD--LT	608	
Db	548	QSMK-----FKGNGIKLNDFDES-----FGWNYNSLNGYFFTKGVKASLGRVTV	592	
QY	609	VFGDKTHQKV--YQG-----NIYRPFIKKSVLRGVAKILGYGN-NLPFYENFVAGYGS	660	
Db	593	IPGSDNKKYKLSADVQGYFPLDRDRHWVSAKASAGYAN-GFGNKRLLPFYQTTTAGIGS	651	

```

Qy 661 VRGVDQSLSLGRPRSOAYLTARGOOT---TLGEVVGGNALATFGSEILPLPFKGDWI-DQ 711
   :||: :|||: | | | | :|||: | | | | | | | | | | | | | | | | | | | | | |
Db 652 LRGFAYSIGPNA---IYAEGHNGTFNKISSDVIGGNAITTAAGELIVPTPFVSFKSQNT 708
   :||: :|||: | | | | :|||: | | | | | | | | | | | | | | | | | | | | | |
Qy 717 VRPVFTEGGOVEDTT-GMDKQTIDLTFQKDPOATAEQNKAANRPLLTTQDKOLRYSGY 775
   || :|||: :|||: | | | | | | | | | | | | | | | | | | | | | | | |
Db 709 VRTSLFVDAASVMWTKWSKNGLSKVLKD-----LPDYGKSSIRASTGV 755
   || :|||: :|||: | | | | | | | | | | | | | | | | | | | | | | | |
Qy 776 GATWTPTGIPLSISYAKPLNKKNDQTDTVQFGTGSVF 813
   | :|||: | | | | | | | | | | | | | | | | | | | | | | | |
Db 756 GFQMOSPGLPVFSYAKPIKKYENDVDEQFQFSIGGSF 793
   | :|||: | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 13
US-08-433-522A-8
; Sequence 8, Application US/08433522A
; Patent No. 6013514
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433,522A
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-434 MIS:jfb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 792 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-433-522A-8

```

[illegible]

QY	202	DVLAIKDNKINPLNSKADRYTOBKLYTSLNRAKYLNAGVFREIKDAKLINEDKNRIF	261
Db	193	EQMELQDPSWKU--WGNKKEGAFQKDLQAIRDYILNNGYAKAQITKAQVQLNDEKTKVN	251
QY	262	VEISLHEGEQYRFQOTFGNLTYYTOAELEALLK--FKAEEGFSQAMLEOTTNNISTKFGD	320
Db	252	VTIDVNEGLQDLYRSARIIGNLGGMSAELEPLLSALHLNDTFRSDIADVENAIRAKLGE	311
QY	321	DGYYIAQIRPVTRINDESTVDVEYIDPVHPVYPRINFTGNFKTQDEVIREMRQLEG	380
Db	312	RGVNTVNSVPDQDANKTLAITFVVDAGRRLTQHLRFEGTYSADSTLROEMRQEG	371
QY	381	ALASNOIKOLSRALMRTEFFKHVTVDTR--PVPNSPDQVDVNFVVEEQSPSGSSTIAAGY	438
Db	372	TWYNSQLVELGKIRLORTGFFE--TVENRIDPINGSNDEVDVYVKERTGCSINFGICY	429
QY	439	SQSGGVTFQEDVSYQNNFMGTGKHVNASFBSRSETREVYSLGMTNPFTVNGVQSQSLSYIR	498
Db	430	GTESGISYQASVKQDNFLTGAASVAGTKNDYGTGVSNLGYTPEYPTKDGV--SLGGNVF	487
QY	499	KTYIDNKK--NTSNVYLDVSYGSGLSGYPIDENORISFGL----NADNTKLHGGRFMI	550
Db	488	FENYDNSKSDTSNYKRTYIGSVNVLGFPVPNNNSYVGLGHTYKNISNFALEYNRNLYI	547
QY	551	SNYKQLMADGGKIQVDNNGIPDKFHDYTTYNAILGNWYSSLDRPVFPQGMGSHVD--LT	608
Db	548	QSMK-----FKGNIKTNDDFS-----FGWYNSLARGFYFPTKGVKASLGGRVT	592
QY	609	VFGDGKTHOKVYVQGNYYIRPF-----IKKSVLRGYAKLGYGN-NLPYENFYAGGYS	660
Db	593	IPGSDNNYYKLSADVGQFPLDRDHLWVYSAKASAGYAN--GFENKRLPYQYITAGTGS	651
QY	661	VRGYQDSSLGPRSQAYLTARRGOQTTL----GEWYGGNALTATGSELPLPLPFKGDWI--D	715
Db	652	LRFAYGSGISGPN-----TYQGNKNKFNKISSDVIIGNAIAIASAELIPTPFVSDKSON	706
QY	716	QVBPVPIFEGGOVEDTT--GMDKOTIDLTCFKDQPAEQAQNAKRNAPLLTQDKQLRYSAG	774
Db	707	TVFTSLFVDASVWNTKWSKDNKNGLESNVLD-----LDPYGKSSRTRASTG	753
QY	775	VGATWYTPIGPLSISYAKPLNKKQNDQTDTVQFGISVF	813
Db	754	VGFQWQSPGSPVVFYAKPIKPYKENDVDQEQFSGISGSF	792

RESULT 14

```

US-09-135-166-8
: Sequence 8, Application US/09135166
: Patent No. 6083743
:
: GENERAL INFORMATION:
:
: APPLICANT: CHONG, Pele
: APPLICANT: THOMAS, Wayne
: APPLICANT: YANG, Yan Ping
: APPLICANT: LOOSMORE, Sheena
: APPLICANT: SIA, Dwo Yuan Charles
: APPLICANT: KLEIN, Michel
:
: TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
:
: NUMBER OF SEQUENCES: 55
: CORRESPONDENCE ADDRESS:
: ADDRESSER: Sim & McBurney
: STREET: 6TH Floor, 330 University Avenue
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5G 1R7
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/135.166

```

Query Match	24.8%;	Score 1042;	DB 4;	Length 792;
Best Local Similarity	31.4%;	Pred. No. 31e-75;		
Matches 257;	Conservative 159;	Mismatches 335;	Indels 68;	Gaps 20;
QY	23	STHAQAADFMANDITITGLQRVITIESLQSVLPFLGQVSNQLADGVKALYATGNFSDV	82	
		:                     : : :           : :		
Db	14	TTTTFAAPFVAKDIRDVGVDGLEQIQIRASLPYRAGQRTDNDVANIVRSLSVSGREDVV	73	
QY	83	QVYHQEGRI-IYQVTERPLIAEINFEGRNLRIPKEGLEGLKNAGLAVGQPLKQATVOMIE	141	
		:         :   : : : :           : :		
Db	74	KA-HQEGDVLVSVVAKSIISDVKIKGNSIIPPEALKQNLNDANGFKVGGDILIREKLENEA	132	
QY	142	TELTNOYISQYZYNTETTVKQTMLDGGRVKNKLDFTFAEGKPARVVDINIIGNHFESADLI	201	
		:         :         : : :		



---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 09:55:54 ; Search time 69.95 seconds  
(without alignments)  
1116.807 Million cell updates/sec

Title: US-09-701-711-4  
Perfect score: 4202  
Sequence: 1 MRNSYFKGFQVSAMTAVMM.....LNKKQNDQDTDTQFQIGSVF 813

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues  
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_71.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1451.5	34.5	797	2 H83190	probable outer mem
2	1262	30.0	795	2 AC0129	probable surface a
3	1218.5	29.0	803	2 A80530	outer membrane pro
4	1210	28.8	810	2 A84742	hypothetical prote
5	1210	28.8	810	2 C90651	hypothetical prote
6	1210	28.8	810	2 C85502	hypothetical prote
7	1203	28.6	803	2 B82099	surface antigen VC
8	1172.5	27.9	784	2 E82731	outer membrane ant
9	1115	26.5	797	2 G81228	outer membrane pro
10	1111	26.4	797	2 D82000	outer membrane pro
11	1053.5	25.1	808	2 F64102	protective surface
12	1052.5	25.0	797	2 JC4078	protective surface
13	754	17.9	768	2 D71726	outer membrane pro
14	734	17.5	774	2 D97527	ompl protein presu
15	734	17.5	774	2 A82746	group 1 outer memb
16	727.5	17.3	769	2 F87486	outer membrane pro
17	697	16.6	768	2 B97725	outer membrane pro
18	680.5	16.2	781	2 A83355	outer membrane pro
19	641	15.3	617	2 H84957	hypothetical prote
20	568.5	13.5	778	2 C70412	outer membrane pro
21	515	12.3	739	2 A81430	outer membrane pro
22	434	10.3	916	2 G64601	protective surface
23	429	10.2	906	2 F71910	probable outer mem
24	397.5	9.5	792	2 H81693	outer membrane pro
25	396.5	9.4	792	2 B71539	probable omp85 ana
26	395	9.4	790	2 D86528	omp85 analog [imp
27	395	9.4	790	2 D72094	probable outer mem
28	356.5	8.5	853	2 A71339	probable outer mem
29	325	7.7	821	2 B70199	outer membrane pro

## ALIGNMENTS

## RESULT 1

H83190  
probable outer membrane protein PA3648 [imported] - Pseudomonas aeruginosa (strain PA  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: H83190  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa  
A:Reference number: A82950; MUID:20437337  
A:Accession: H83190  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-797 <STO>  
A:Cross-references: GB:AE004784; GB:AE004784; PIDN:g9949799; PIDN:AAG07036.1; GSPDB:GN  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA3648  
C:Superfamily: protective surface antigen D-15

Query Match 34.5%; Score 1451.5; DB 2; Length 797;  
Best Local Similarity 36.9%; Pred. No. 4.2e-76;  
Matches 308; Conservative 163; Mismatches 297; Indels 67; Gaps 13;

QY	7	KGFQVSAMTAVMMVMVMASTHAQAADFMANDITITGLQRTVIESLQSVLPFRLGQVVSQNL	66
DB	2	KRFLPAL-LSALMIAEVHAES--FTVSDIRVNGLQVSAGSVFAALPLNVGETIDQAL	58
QY	67	AGCVKALYATGNFSDVQVYHQEGRIIYQVTERPLIAEINFEGRNLTPKEGLQGLKNAGL	126
DB	59	VOATRSLFKTFGFDIQLGRDGNVLVTVVVERPSISIEIEGNAKSKEDLLKLGKOSGL	118
QY	127	AVGOLPKQATVOMIEFTLNQYISQGYVNTETVKTQMTLDGNNVKLDMTFAECKPARVVD	186
DB	119	AGEIFORATLEGRVNELOQYVQAQGRYSAEINAEVPOPNRVALKININETGVAASH	178
QY	187	INIIGNQHSADLIDVLAIK-DNKINPLSKADRYTQEKLVTSLENLRAKYLNAAGVRF	245
DB	179	INVGVNTVSEEDLTDLFELKTTNWLSPFKNDKDYAREKLSGDLRLRSYLLDORGYNMD	238
QY	246	IKDAKLNINEDKNRIVEISLHEGEQYRFGQTOFLGNLTYTQAEALAKFAEFGSQA	305
DB	239	IASTQVSIITPKKHVYITVNIINEGEKYITRDVKLTGDKVPPEEVEVKRLLLVQKQVFSRK	298
QY	306	MLEQNTNNISTKFGDGYVYAOIRPVTRINDESRVDVEYIDPVHPVVYRRINFTGNFK	365
DB	299	VMTTSDLLITRLNGNEGYTFANVNGVPEAHDDDKTQVSVTFVDPGKRAYNRINFRGNTK	358
QY	366	TQDEVLRREMRQLEGALASNQIKQLSRARLMRTGFFKFKHTVTDTRPVPNSPDQDVNFVVE	425

```
Db 359 TEDEVLRREMRQEGGWASTYLLDQSKARLERLGYEKEVNVETPAVPGTDDQVDVNSYVE 418
QY 426 EQPSGSSSTAAGYSQSGGVTFQFDVSONNFMGTGKHVNASFSESRSEPREVYSLGTMNPPYT 485
Db 419 EQPSGSSITASGAQAGLILGSSISQNNFLGTGNKVSIGLRKSEYQTYRNFQFVDPYT 478
QY 486 VNGVSOSLSGYRKTXYD--NKNISNYSVLDSSYSGSLSYGPIDENQISFGLNADNTKLH 543
Db 479 VDGVSLSGYNFAFYKTDYDELDDVASYSVNSLGAAGSICGPISETSRLTYGLSVQDQID 538
QY 544 GREFMGISNVKQLMADGGKIQDNNGIPDFKHDTYTYNAILGWNYSLLDRPVPFTQGMWH 603
Db 539 TGRYT-VDEIYDFLKEG-----DN-----FTNFKASICWSESTLNGKGLATRGHSQ 584
QY 604 SYDL--TVGFGDKTHQVYVYQNIYRPFTHKSVLRGYAKLGYGN-----NLPEYENFA 655
Db 585 SLTLETLPGSLSFYKIDYRGQVAPLTDNTMRFHFLGYGDSYSTERLPFTYENYA 644
QY 656 GYGVSVRGYDQSSLGPRSQAYLTARRGQQTTL-----GEYVGGNALAT 698
Db 645 GGFNSVRGFKDGLGRSTPSV-ARNPDGTPMKNOGPDCKGRYTDQDQDPEAFGGNIIIT 703
QY 699 FGSEILPLPFPGKDWIDQVRPFIETEGGVFTTGMKQOTIDLTKPQDQAEONAKAA 758
Db 704 GCAELLPLPFVKDQ-RQLRTVLFWDVGSFTDCTPTTTTNCQGIK-----749
QY 759 NRPLITQDKOLRYSGAGVATWPTPGPLSISYAKPLNKKQNDQTDVQFQIGSVF 813
Db 750 -----TDNLASSVGVGLTWITALGPLSLFSLATPIKPDNAETQVQFSLGQTF 797

RESULT 2
AC0129
probable surface antigen YP01052 [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C:Accession: AC0129
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, P.; Skelton, M.; Simmonds, M.; Skelton, J.; Wain, J.; Church
th, T.; Connerton, P.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farr
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB00001; MUID:21470413; PMID:11586360
A:Accession: AC0129
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-795 <KOR>
A:Cross-references: GB:AL590842; PIDN:CAC89894.1; PID:g15979119; GSPDB:GN00175
C:Genetics:
C:Gene: YP01052
C:Superfamily: protective surface antigen D-15

Query Match 30.0%; Score 1262; DB 2; Length 795;
Best Local Similarity 35.7%; Pred. No. 3.8e-65;
Matches 292; Conservative 152; Mismatches 329; Indels 44; Gaps 14;

QY 14 MTMAVMMVMSTHAQAD-EMANDITITGLQRTVIESLQSVLPRLGQVSVENQLADGVKA 72
Db 6 LLIASLLFGSATYVGADGVVNDIHFEGLQAVGAALLNMPVRVGVDTVSDDDIGKTIRA 65
QY 73 LYATGNSDQVYHQEGRIYQVTERPIAIEINFEGRNLRIPKEGLOBLKNAGLGOPL 132
Db 66 LFATGNFEDVRVLDGNTLVQVKERTIASITFGSGNKAVKEDMLKNLEASGVVRGEAL 125
QY 133 KOATVQMIEITELNQYISQGYNYFETIVTKOTMLDGNRKLDMTFAEGKPARVVDINIIGN 192
Db 126 DRTTISNIEKLEDFYISVGKYSASVAVVTPLRNRRVLDKLVTETGVSAKIQIINIVG 185
QY 193 QHFSADALIDLVAIKD-----NKNINPLSKADR-YTOEKLVTSLENRAKVLNAGVRFPEI 246
Db 186 HSTTDELISRFQIRDEVPPWNV-----GDRYKQKQKLAGDLETURSFYLDGRGARFNI 240
```

```
QY 247 KDAKLINEDKNRIFVEISLHGEQYRFQQTQFLGNLTYTQAEALLKFAEEGFSQAM 306
Db 241 DSTQVSLTPDKKGIYITINITEGPQFKLNSVSVAGLHQAEBKLTKEPCEGLFNGSK 300
QY 307 LEQTTNNISTKFGDDGYIAQIIRPVTRINDESTRVDVVEYIIDPVHPVYVRRINFTGNFKT 366
Db 301 VTRMEDDITKMLGRYGAYPRVVTQPEINDDKTVKLHINVDAGNRFYVYRHRFRFEGNPTS 360
QY 367 QDEVLRREMRQLEGALASNOKIQLSRARLMRTGFFKHVTVDTRPVNSPDQVDVNFVVEE 426
Db 361 KDSVLRREMRQMEGAWLGNDOVEACKERLNLRLGYETVDVETQVRPGADLVDTYKYKE 420
QY 427 QPSGSSSTAAGYSQSGGVTFQFDVSONNFMGTGKHVNASFSESRSEPREVYSLGTMNPPYT 486
Db 421 RNTGSLNFGIGYGTSGSVFQVQDDNMLGTGNTGVINGTKNDYQTYAEFTLMDPYETV 480
QY 487 NGVSOSLSGYRKTXYDNKNISNYSVLDSSYSGSLSYGPIDENQISFGLNADNTKLHGR 546
Db 481 DGVSLLGGRIFYNDFKADNADLSGYTNSYGADGTGLGFFPINENNSLRVGGVYVHNDL-SDM 539
QY 547 FMGISNVKQLMADGGKIQDNNGIPDFKHDTYTYNAILGWNYSLLDRPVPFTQGMHSDV 606
Db 540 LPQVAMWRYLESVGERPGYD--GREGFTDDFTLN--LGWTYNNLDRGFFFTSGVKSSVN 595
QY 607 -LTVGFQDKTHQVYVYQNIYRPF--IKKSVLRGYAKLGYGN-----NLPEYENFYAGG 657
Db 596 TKITVPGSDNEFYKVTFTSAYPLNEDRSWLLGRGLGYGDSGKSEMPFENFYAGG 655
QY 658 YGVSVRGYDQSSLGPRSQAYLTARRGQQTTLGCVVGGNALATFGSELILPLPKGD-WIDQ 716
Db 656 SSTVRGFRSNNIGPKA-AYYANGATVNTSDAVGGNAMAVASIELITPTTFISEKYSNS 714
QY 717 VRPFIETEGGVFTTGMKQOTIDLTKPQDQAEONAKAANRPLITQDKOLRYSGAGV 776
Db 715 VRTSIFIDSGVFWDTNW-----ENTAK--TRAAGIPDYCKASNIRVSAGVA 758
QY 777 ATWTPICPLSISYAKPLNKKQNDQTDVQFQIGSVF 813
Db 759 LQWMSPLGLPVFSYAKPVKDYEGDKSEQFQFNIGKTW 795

RESULT 3
AB0530
outer membrane protein precursor yaef [imported] - Salmonella enterica subsp. enteric
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
C:Accession: AB0530
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Ouail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se
A:Reference number: AB0502; PMID:11677608
A:Accession: AB0530
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-803 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD08682.1; PID:g16501505; GSPDB:GN00176
C:Genetics:
A:Gene: yaef
C:Superfamily: protective surface antigen D-15

Query Match 29.0%; Score 1218.5; DB 2; Length 803;
Best Local Similarity 33.9%; Pred. No. 1.2e-62;
Matches 285; Conservative 150; Mismatches 327; Indels 79; Gaps 15;

QY 14 MTMAVMMVMSTHAQADFMANDITITGLQRTVIESLQSVLPRLGQVSVENQLA 67
Db 1 MAMKKLLIASLLFSSSATYVGAEGFVVKDIHFEGLQAVGAALLSMFVTRGTVDNEDIS 60
```

Qy	68	DGVKALYATGNFSDVQVYVHQBGRILIQVTERPLIAEINTEGRLIPKELQBLGNAGLA	127
Db	61	NTIRALEATGNFEDVRLVDGNLTLLVQVKERPTIASITPSGNKSVKDDMLKQMLEASGYR	120
Qy	128	VGQPLKQATQVMIETELTNOYISQYGYNTETITVKQTMDCNRVKLDMTAEGRPARVDI	187
Db	121	VGESLDRITLSDIEKLEDFYYSVGKYSASVKAIVTPLPNRNVYDLKLIVQEGVSAKIQI	180
Qy	188	NIIGNQHFSADLIDVLAIKD-----NKINPLSKADR-VTQEKLVTSLENLAKYINAGF	241
Db	181	NIVGNHAFSTEELISHPOLDFEPPWNVV-----GDRKQKQKLAGDLETLRSYVLDRCY	235
Qy	242	VRPEIKDAKLINEDKNRIEVEISLHEGEQYREFQGTQFLGNLTNTCAEFAELILKPKAE	301
Db	236	ARFNIDSTQVSLPDKGIIYIWNTEGDYKLSGVQVSGNLGASHAEIEKLTKIPGGU	295
Qy	302	FSQAMLEQTTNNISTKFGDDGYAAQIRPVTRINDESRTVDVEYYIDPVHPVYVRRINET	361
Db	296	YNGTKVTKMEDDIKKILGRYGYAYPRVQSOPETINDAKTVKLRNVVDAGNRFYVRKIRE	355
Qy	362	GNFTQDEVILRRMRQLEGALASNOKIQLSRALRMTGTFKFKHYVDTRPVPNSPDQVDV	421
Db	356	GNDSKDSVILRRMRQEGAWLGDSDLVDQGERLNLGRGFETVDTQTRVPSPDQVDV	415
Qy	422	FVVEEQPSGSTIAAGYSQSGGYTFQFVSQNNFMGTGKHVNAFSSRSRETVYSLGTMN	481
Db	416	YKVKERTGSEFNGIGYGTESGSYFQAGVQDNNLGTGYSVGLNGTKNDYQYSELSVTN	475
Qy	482	PYFTVNGVSQSLGYRYKTKYDKNNKINSYVLDYSGSLSYGYPIDENQRISFGLNADNTK	541
Db	476	PYFTVDCVSLGGRIFYNDFFADADLSDYTNKSYGTDVTLGFPINRYNTLRAGLGYVHNK	535
Qy	542	LHGCRFMGISNVK-QLMADGGKIQVDNNGTIPDKFKHYDTYTNATILGNYSLSLDRPVEPT	600
Db	536	L-----SNMQPQIAMDRIYLSMGDPDASDFAADDFTFN--YGWTYINLDRGTFPTDG	585
Qy	601	--NSHVSVDLVGVGDKTHQKVYVQGYNIYRPFK--KSVLRGYAKLGYGNL-----PFYE	651
Db	586	SRVNLTKGVITPGSDNEYKYSVLDATYVPIDMDHKWVVLGRTRWYGDGLGKKEMPFYE	645
Qy	652	NFVAGGYGVRYGDDSSLGPRSQAYLTARGQOTT-----LGEVVGGN	694
Db	646	NFTAGGSSTVRGQSNTIGPKA-----VYKNGAHTSWDDDDVEDCTQESGCKSDDAVGN	701
Qy	695	ALATFGSELILPLPFKGD-WIDQVRPVIFTEGGQVDFTT--GMDKQITDLTKFDKQATAE	752
Db	702	AMAVASLEFITPPISEKYANSVRTSFFWDMGTVDNTNWDPSASDPDYSDP-----	756
Qy	753	ONAKAANRPLLTDQKOLRYSAGVAGTATWYTPIGPLSTISYAKPLNKKONDQTDVTFQIGSV	812
Db	757	-----GNIRMSAGIALQMLSPGLPVFSYAQPFKKYDGDRAEQFQFNIGKT	802
Qy	813	F 813	
Db	803	W 803	
RESULT	4		
A64742		hypothetical protein b0177 - Escherichia coli	
C:Species:		Escherichia coli	
C:Date:		12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 29-Sep-1997	
C:Accession:		A64742	
R:Blattner,		F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;	
A.;		Rose, D.J.; Mau, B.; Shao, Y.	
Science		277, 1453-1462, 1997	
A:Title:		The complete genome sequence of Escherichia coli K-12.	
A:Reference		number: A64720; PMID:97426617	
Accession:		A64742	
A>Status:		nucleic acid sequence not shown; translation not shown	
A:Molecule		type: DNA	
A:Residues:		1-810 <BLAT>	

A:Cross-references: GB:AE000127; GB:U00096; NID:gl1786370; PIDN:AAC73288.1; PID:gl178663  
A:Experimental source: strain K-12, substrain MGI655  
A:Superfamily: protective surface antigen D-15

Query Match	28.8%;	Score 1210;	DB 2;	Length 810;
Best Local Similarity	34.0%;	Pred. No. 3.9e-62;		
Matches 287;	Conservative 149;	Mismatches 330;	Indels 78;	Gaps 15;

Qy	14	MTAMVMVMS-----THAQADFMANDITITGLQRTVETIESLSQVLPFRGVQVSVENQLA	67
Db	1	MAKKLLIASLLFSATVYGAEGFVKDIHFEGLQVAVGAALLSMKQNLASGV	60
Qy	68	DGVKALATGNFSDVQVYVHQBGRIIYQVTERPLTAETFEGRNLIPKEGLQGLKNAGLA	127
Db	61	NTIRALFATGNFEDRVLRDGDTHLLVQVKERPTTASTTFSGNKSVDKMLKQNLASGV	120
Qy	128	VQPLKQATVQMIETEITNQYISOGYVNTETITVKTQMLDGNRVKLDMTFAEGKPARVVDI	187
Db	121	VGESLDRTTIAIDIEKLEGFYISVGKYSASVKAVVVPLPRNRVDLKLVFQEGVSAEIQOI	180
Qy	188	NIIGNQHFSDADLIDVLAIKD-----KNINPLSKADR-YTQEKLVTSLENLRKAYLNAGF	241
Db	181	NIVNGHAFPTDDELISHFQLRDVEPMMWV-----GDRKYQOKIAGLDLETILRSYLDRGY	235
Qy	242	VRFEIKDAKLINDEKNNRIFVEISLHGEQYRFQGTQFLGNLTATQAEALLKFKAEAG	301
Db	236	ARFNIDSTQVSLTPDKKGIYVTVNITGSDQYKLSGVEVSGNLACHSAEIQLTKEIPGEL	295
Qy	302	FSQAMLEQTTNNISTKFGDDGYYIAQIRPVTRINDESRITVDVEYYIDPVHPVYVRRINFT	361
Db	296	YNGTKVTKMEDDIKLLGRGYAYPRVQSPEINADAKTVKLRVNDVAGNRFVVRKIRFE	355
Qy	362	GNFTQDEVLRREMRQLEGLASNOKITQLSRARLMRTGFEFKHVTVDTRPVNPSDQVDVN	421
Db	356	GNDSKDAVLRREMRQEGAWLGSDLVDQCKERLNRLGFFETVDTDTQRPVPGSDQVDVV	415
Qy	422	PVVEEQPGSGSTIAAGYSQSGVTTQFDVSONNPMGTGKHVNASFRRSETREYVSLGMTN	481
Db	416	YKVERNTGPNFGIGYGTESGSFQAGVQDDNMLGTGYAVGINGTKNDYTQVAAELSVTN	475
Qy	482	PYFTVNGVSQSLGYRKYVDNKNINSVYLDYSVGSLSYGYPIDENQRISFGLNADNTK	541
Db	476	PYFTVDGVSGLGRIFYNDFOADDADLSYTNKSYGTDVTLGFPINENSLRAGLGYVHNS	535
Qy	542	LHGGRFMGLSNVKOLMA-----DGGKIQVDNNGIPDKHDTYTNAILGWNYSLSLRPV	595
Db	536	L-----SNMQQVAMRWRLYSKGEHPSTSDODNSFKTDDFTFN--YGYTNYKLDRGY	585
Qy	596	PPTGG--MHSIVDLTVFGDKTHQKVYVQGNIRPF--IKKSVLRGYAKILGYGNL----	647
Db	586	FPTDGSRVNLTKGVTIPGSDNEYKYKVLDTATYVIPIDDDHKVYVGLTRWGYDGLGGKE	645
Qy	648	-PVENFYAGCYGVRGYDQSSILGPRS-----QA-----YLTARRGOOTTL--GEV	690
Db	646	MPFENFYAGGSVTRVFGQSNITGPKAVYEPHQNSNTPDPTDECATQCAKDLCKSDDA	705
Qy	691	VGGNALTATFGSELILPFFKGD-WIDQVRPIFTEGGQVFDTTGMCKQTIDLTQFKDPQA	749
Db	706	VGGNAMAVALSEFITPTPTFISDKYANSVRTSFFWDMCTVMDTNWSSQYSQYSDYSDP--	763
Qy	750	TAEONAKAANRPLLTDQKQLRYSGAGCATWYTPIGPLISYAPKPLNKKQNDQTDYVQFOI	809
Db	764	-----SNIRMSAGIALQWMSPLGLPLVFYSYAQPFKYDGDKAEOQFNI	806
Qy	810	GSVF	813
Db	807	GKTW	810
RESULT	5		
C90651			

RESULT 5  
C90651  
hypothesis

C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
C:Accession: C90651  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: C90651  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-810 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BAB33602.1; PID:gl3359635; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RIMD 0509952  
C:Genetics:  
A:Gene: ECs0179  
C:Superfamily: protective surface antigen D-15

Query Match 28.8%; Score 1210; DB 2; Length 810;  
Best Local Similarity 34.0%; Pred. No. 3.9e-62;  
Matches 287; Conservative 149; Mismatches 330; Indels 78; Gaps 15;

QY 14 MTMAVMVMMS-----THAQAAADFMANDITITGLQRTVIESLQSVLPFRLGQVVSQSLA 67  
DB 1 MAMKLLIASLLFSSATVYGAEGFVVKDIHFEGQLQRVAVGAALLSPVRTGDTVNDSDIS 60  
QY 68 DGKALYATGNFSDVOVYHQBGRITTYQVTERPLIAINFEGRNLPKEGLQGLKKNAGLA 127  
DB 61 NTRALFATGNFEDVRVLRDGTLLVQVAKERTPIASITFTSGNKSVDKMLKQNLASGVR 120  
QY 128 VGQPLKQATVQMIETELTNYISQGYNTETITVKQTMLDGNRVKLDMTFAEGKPARVVDI 187  
DB 121 VGESLDRITTIADIEKLEDFYYSVGKYSASVAVTPLPRNRRVDLKLVFQEGVSAREIQOI 180  
QY 188 NIIGNHFSADLIDVLAIKD-----NKINPLSKADR-YTQEKLVTSLENLRAKYLNAGF 241  
DB 181 NIVGNHAFTTDELISHFQLRDEVPWNV-----GDRKYQKQKLAGDLETLSRYLDGY 235  
QY 242 VRPEIKDAKINEDKNRIFVEISLHQBGRITTYQVTERPLIAINFEGRNLPKEGLQGLKKNAGLA 301  
DB 236 ARNIDSTQVSLTPDKKGIYVTVNITEGQYKLSGVEVSGNLASGAEIEQLTKIEPGL 295  
QY 302 FSQAMLEQTTNNISTKFGDDGYGYAQRVTRINDSRVTDVEYYIDPVPVVRVRRINF 361  
DB 296 YNGTKVTKMEDDIKLLGRYGAYPRVQSMPEINDADTKVLRVNDAGNRFVVRKIRFE 355  
QY 362 GNFKTOEVLRRERMOLEGASLNQIKLSRARLMTGFFKHVTVTRPVNSPDQVDVN 421  
DB 356 GNDTSKDAVLRERMRQMEGAWLGSDLVDQGERLNLGFFETVDTQTRVPGSPDQVDV 415  
QY 422 FVVEEQPSGSSTIAAGYSQSGVTFQFDYSQNNFMGTGKHVNASFSRSETREYVSLGTMN 481  
DB 416 YKVKERTGTSFNGIGYTESGVSFOAQVQDNLGTYAVGKNGTKNDYQTVAEISLVN 475  
QY 482 PYFTVNGVSQSLSGYRKYTKYDNKNISNVLDSYSGSLSYGYPIDENQRISFGLNADNTK 541  
DB 476 PYFTVNGVSLGRLFYNDQADDADLSDTNKSYGTDVTLGFPINEYNSLRAGLVVHNS 535  
QY 542 LHGGRFWSINVKQLMA-----DGKIQVDNNGIPDEKHYTYNAILGNWYSSLDPRV 595  
DB 536 L-----SNMQPQVAMRWLYLSMGHEPSTSDQNSFKTDDFTFN--YGWTYKRLDGY 585  
QY 596 FPTQG--MHSVSDTLVGFQDKTHQVYQYNIYRPF--IKKSVLRGYAKLGYNNI----- 647  
DB 586 FPTDGSRVNLTKVTPGSDNEYKYTLDTATVPIDDDKHVVVGLTRNGYGDGLGGKE 645  
QY 648 -PPYENFYAGGVSRYGQSSIGLPRS-----QA-----YLTARRGQQTFL-----GEV 690  
DB 646 MPYENFYAGGVSRYGQSSIGLPRS-----QA-----YLTARRGQQTFL-----GEV 690  
QY 691 VGNALATGSELILPLPKGD-WIDQVRPVIIEGQVDFDTGMDKOTIDLTQPKDPOA 749

DB 706 VGNAMAVASLEPITPTPFISDKYANSVRTSFWMGTWMDTNDSSQYSGVPDYSDBP-- 763  
QY 750 TAEQNAKAANRPLLTQDKOLRYSGAGVATWYTPIGPLTSISYAKPLNKKONDOQDTDTVOFOI 809  
DB 764 -----SNIRMSAGIALOWMSPLGFLVFSYAQPFKKYDGDCAEAEQFQFNI 806  
QY 810 GSVF 813  
DB 807 GKTW 810

RESULT 6  
C85502  
hypothetical protein yaeF [imported] - Escherichia coli (strain O157:H7, substrain ED  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C:Accession: C85502  
R:Perina, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May  
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouasis, K.; Apoda  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: C85502  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-810 <STO>  
A:Cross-references: GB:AE005174; NID:gl2512902; PIDN:AAG54479.1; GSPDB:GN00145; UWGP:  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: yaeF  
C:Superfamily: protective surface antigen D-15

Query Match 28.8%; Score 1210; DB 2; Length 810;  
Best Local Similarity 34.0%; Pred. No. 3.9e-62;  
Matches 287; Conservative 149; Mismatches 330; Indels 78; Gaps 15;

QY 14 MTMAVMVMMS-----THAQAAADFMANDITITGLQRTVIESLQSVLPFRLGQVVSQSLA 67  
DB 1 MAMKLLIASLLFSSATVYGAEGFVVKDIHFEGQLQRVAVGAALLSPVRTGDTVNDSDIS 60  
QY 68 DGKALYATGNFSDVOVYHQBGRITTYQVTERPLIAINFEGRNLPKEGLQGLKKNAGLA 127  
DB 61 NTRALFATGNFEDVRVLRDGTLLVQVAKERTPIASITFTSGNKSVDKMLKQNLASGVR 120  
QY 128 VGQPLKQATVQMIETELTNYISQGYNTETITVKQTMLDGNRVKLDMTFAEGKPARVVDI 187  
DB 121 VGESLDRITTIADIEKLEDFYYSVGKYSASVAVTPLPRNRRVDLKLVFQEGVSAREIQOI 180  
QY 188 NIIGNHFSADLIDVLAIKD-----NKINPLSKADR-YTQEKLVTSLENLRAKYLNAGF 241  
DB 181 NIVGNHAFTTDELISHFQLRDEVPWNV-----GDRKYQKQKLAGDLETLSRYLDGY 235  
QY 242 VRPEIKDAKINEDKNRIFVEISLHQBGRITTYQVTERPLIAINFEGRNLPKEGLQGLKKNAGLA 301  
DB 236 ARNIDSTQVSLTPDKKGIYVTVNITEGQYKLSGVEVSGNLASGAEIEQLTKIEPGL 295  
QY 302 FSQAMLEQTTNNISTKFGDDGYGYAQRVTRINDSRVTDVEYYIDPVPVVRVRRINF 361  
DB 296 YNGTKVTKMEDDIKLLGRYGAYPRVQSMPEINDADTKVLRVNDAGNRFVVRKIRFE 355  
QY 362 GNFKTOEVLRRERMOLEGASLNQIKLSRARLMTGFFKHVTVTRPVNSPDQVDVN 421  
DB 356 GNDTSKDAVLRERMRQMEGAWLGSDLVDQGERLNLGFFETVDTQTRVPGSPDQVDV 415  
QY 422 FVVEEQPSGSSTIAAGYSQSGVTFQFDYSQNNFMGTGKHVNASFSRSETREYVSLGTMN 481  
DB 416 YKVKERTGTSFNGIGYTESGVSFOAQVQDNLGTYAVGKNGTKNDYQTVAEISLVN 475  
QY 482 PYFTVNGVSQSLSGYRKYTKYDNKNISNVLDSYSGSLSYGYPIDENQRISFGLNADNTK 541  
DB 476 PYFTVNGVSLGRLFYNDQADDADLSDTNKSYGTDVTLGFPINEYNSLRAGLVVHNS 535

```
QY 542 LHGGFMGLSNVKKQLMA-----DGGKIQVDNNGIPDFKHDTYTTNAILGWNYSLLDRPV 595
Db 536 L-----SNMQPQVAMWRYLSMGHEPSTSDQNSFKTDDFTFN--YGTWYNKILDRGY 585
QY 596 RPTOG--MHSVVDLNVGFGDKTHQVYVYOGNLYRPF--IKKSVLRGYAKLGYGNL----647
Db 586 FPDGSRNLTOKVTPGSDNEIYKVTLDATYVPLDDHKKVWVLGTRRWGTYGDLGGKE 645
QY 648 -PFYENFYAGGSGVRGVDQSSSLGPRS-----QA-----YLTARRQQQTTL---GBV 690
Db 646 MPFYENFYAGGSGTVRGQSNITGPKAVYFPHQASNYDPDYDECATODGAKDLCKSDDA 705
QY 691 VCGNALATFGSELIILPLPFKGD-WIDQVRPVITIEGGQVFDTGMDKQIDTIDTQKPOQA 749
Db 706 VCGNAMAVASLEFITPTPFISDKYANSVTSFFWDMGTWMDTNDSSQSGYPDYSDP--763
QY 750 TAEQNAKAANRLLTQDKOLRYSAGVATWYTPIGPLSTSIAKPLNKKQNDQTDIVOFQI 809
Db 764 -----SNIRMSAGIALQWMSPLGLPVFSYAQPFPKPYDGDRAEQFQFNI 806
QY 810 GSVF 813
Db 807 GKTW 810

RESULT 7
BB2099
surface antigen VC2252 [imported] - Vibrio cholerae (strain Ni961 serogroup O1)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-feb-2001
C:Accession: BB2099
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, B
L, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: AB2035; MUID:20406833
A:Accession: BB2099
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-803 <HE>
A:Cross-references: GB:AE004297; GB:AE003852; NID:g9656810; PIDN:AAF95396.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain Ni961; biotype El Tor
C:Genetics:
A:Gene: VC2252
A:Map position: 1
C:Superfamily: protective surface antigen D-15

Query Match 28.6%; Score 1203; DB 2; Length 803;
Best Local Similarity 33.4%; Pred. No. 9.8e-62;
Matches 278; Conservative 155; Mismatches 328; Indels 72; Gaps 15;

QY 16 MAVVMVMTAAQAD-EMANDITITGLQVVTIESLQSVLPFRLGQVSVENQIADGVKALY 74
Db 8 LATLLTSVSAAGKEFVVQDQIDQLQVLAALLKMPVRGVSQDVANIKALY 67
QY 75 ATGNFSDVQVYHQEGRIVYQVTERPLIAEINFEGRNLIPKEGLQBLKNAGLAVGQPLKQ 134
Db 68 SSGNFEDVKVLRDGTNLVQVKERTIASVSPGSKAIEEQQLKNLEASSIRVGEALDR 127
QY 135 ATVQMIETELTNQYTSQGYNTIEITVKOTMLDGNRVKLDMTFAEGKPARVVDINIGNQH 194
Db 128 TTLSNIERGLEDFYYSVGKYNATVAVVTPPLPRNRADLKFVFTGEGVSAKIQINFTGNQV 187
QY 195 FSDADLIDVIAKDKNI--NPLSKADRYTOEKLVTSLNLRKYLKNAGVFRFEIKDAKLN 252
Db 188 FSDEELISRFNLVNDVAMNVLAD--DKYKKQVLGADIEALRTYYLDRGYLKQVDDSTQVA 246
QY 253 INEDKNRIFVEISLHEGQYRGOTQFLGNLTYYTOAELEALLKFAEFGSOAMLEOTTN 312
Db 247 ISPDKKGVYITLNLNHEGEPYTSKVQFRGELMGKAEFTSLIPFEIGETINGSVATRL EE 306
```

```
QY 313 NISTKFGDDGYIAQIRPVTRINDESRTVDVVEYYIDPVHVPVYRRINFTGNFTQDEVLR 372
Db 307 SVKKVLGEGSAYAPQVRTIPEFDEKQOQSLVVHVEAGKRVYVRDIRFVGNSTRDEVLR 366
QY 373 REMRQLEGALASNOKLTQLSPARLMRTGFFKHVTVDRPVPNSPDQVDVNFVVEQPSGSS 432
Db 367 REMRQEGEGLNSKDIETGKTRLNRLGFFETVEQVTRVPGSEDDQVDLVYSVEKANSNV 426
QY 433 TTAAGYSQGGVTTFQFDVSONNFMGTGKHVNASFSRSETREVYSLGNTNPFYTVNGVSQS 492
Db 427 NFGVGYGTESGVSFQVGLQDNFLGSGNRVGNAMINDYQKNLTLEYRDPYWNLDGVSIG 486
QY 493 LSGYRKTKYDNKNISNYLDVSYGSLSYGYPIDENQIRISFGLNADNTKLHGFRFMGIN 552
Db 487 GKVFYNOFEASEAGIVDYTNESYGTSLTWGFPFDELNRFEFGIGYTHNKI--GNLTPLYLQ 544
QY 553 VKOLMA-----DGGKIQVDNNGIPDFKHDTYTTNAILGWNYSLLDRPVFPPTQGHSHS- 604
Db 545 VENFLAAQASNIDSGGNLLTDD-----FDINLSWTNRNLSNYSFETAG-NHQH 591
QY 605 --VDLTVGFGDKTHQVYVYOGNLYRPFIRKS-----VLRGYAKLGYGN--LP 648
Db 592 AFYKMTVPGSDAQYFKLQYDVRYQYFPLTKKHEFTLLRG--RLGYNGYQOTDGNLFP 649
QY 649 FYENFYAGGSGVRGYDQSSSLGPRS--QAYLTARRGOQTTLGVEVGGNALATFGESELILP 706
Db 650 FYENFYAGGFTSLRGFGNSAGPKAVYRDYSGSNNGSDTATDDSDVGGNALALASVELIVP 709
QY 707 LFPKGDWI--DOVRPVIFIEGGQVFDITGMDKQID----LTFQKPOQATAEQNAKANR 760
Db 710 TPFASEARENRQRTSIFYDMASVMDFEYRGKADYGNQVYYDYSDP-----756
QY 761 PLLTDQKQLRYSAGVATWYTPIGPLSTSIAKPLNKKQNDQTDIVOFQIGSVF 813
Db 757 -----TNRSSYGVALQWVSPMGPLVSLAKPIKKEGDEEFFFTTIGRTF 803

RESULT 8
E82731
outer membrane antigen XF1046 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C:Accession: E82731
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: AB2515; MUID:20365717
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: E82731
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-784 <SIM>
A:Cross-references: GB:AE003941; GB:AE003849; NID:g9105978; PIDN:AAF83856.1; GSPDB:GN
A:Experimental source: strain 9a5c
R:Simpton, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.;
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuranae, E.E.; La
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sava
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF1046
C:Superfamily: protective surface antigen D-15
```

RESULT 9  
G81228  
outer membrane protein Omp85 NMB0182 [Imported] - Neisseria meningitidis (strain MC58 serotype 4) [NCBI]  
C:Species: Neisseria meningitidis  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
C:Accession: G81228  
R:Retellin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; Iri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.  
Science 287, 1809-1815, 2000  
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Vignani, A.; Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.

Query Match	26.5%	Score	1115	DB 2	Length	797	
Best Local Similarity	32.5%	Pred.	No. 1.2e-56				
Matches	266	Conservative	154	Mismatches	361	Indels	38
Gaps	14						
QY	10	QVSAMTMAVMVMSTHAQAADFMANDITITGLORVTIESLQSVLPFRLGQVVSSENQADG	69				
DB	2	KLKQIASALMMLGISPLALADFTTQDTRVEGLQRTSTVFNYLPVKVGOTYNDTHGSAI	61				
QY	70	VKALYATGNFSDVOVYHQEGRITTYQVTERPLIABINFEGRNLIPKEGLQGLKAGLAVG	129				
DB	62	IKSLYATGTFDDVRVETADGQLLTIVTERPTIGSLNITGAKMLQNDAIKKNLESFGLAQS	121				
QY	130	QPLQAQIVQMIETELTNQYISQGYNTETIVQTMPLDGNRVKLDMTFABGKPARVVDINI	189				
DB	122	QYFNQATLNOAVAGLKEEYLGKGLQNTQITPKVTKLARNRVDIDITDEGKSARITDIEF	181				
QY	190	IGNQHFSDADLIDVLAIKDNKI-NPLSKADRYTQEKLVTSLEINRAKYLNAGVFRFEIKD	248				
DB	182	EGNQVYSDRKLMRMSLTGEGITWLTWLRNSQFNQFQKQADWEKVTDFYQNGYDFDRLD	241				
QY	249	AKLNINEDKNRIFVEISLHGEQYRFQOTQPLNLT-YTQAELEALLKFAEGFSQAML	307				
DB	242	TDIQTNEDEKTKQTIKTIIVHEGGRFRWGKVSIEGDTNEVPKAELEKLLTMKPGKWIYERQOM	301				
QY	308	EQTNNISTTFGDDGYYYAOLRPVTRINDESRITVDVEYIDPVHPVYVRRINTEGNFKTO	367				
DB	302	TAVLGEIQNRKMGSGAYAISLSVQPLNATKTVDVFLHIEPGRKIYVNIHITGNKNKR	361				
QY	368	DEVLRMRQLEGALASNQKIQSRLMRGTGFKHHVYVTRPVPNSPDQVDNVFVVEEQ	427				
DB	362	DEVYRRELROMESAPYDTSKLQRSKERVLLGFDNVQFQFVPLAGTDPKVDLNMSLTER	421				
QY	428	PSGSTTAAQYSGSGGYTFQFDYSONFMCTGKHVNASFSSRSETREYVSYSGMNPVETVN	487				
DB	422	STGSLDSAGVQDITGLVMAGSVSDMLFGTGSAAALRASRSKTTLMGSLSTFDPTPTAD	481				
QY	488	GVSSQSLGGYR--KTKYDNKNISNYVLDSYGGSLSYGYPIDENORISFGLNADNTKLHGG	545				
DB	482	GVSLGYDVYKAFDPRKASTSIQYKTTTAGIRMSVPVTEYDRVNFGLVAEHLTVN--	539				
QY	546	RFWGISNVKQLMADGGKIQVDNNGIPDFKHDYTTYNAILGWNYSLLDRPVPPTQGHSHSV	605				
DB	540	---TYNKAPHYADFIKKYKGTGD-TDGSFKWLYKGTGVGMRNKTDTSALWPTRGYLTGV	595				
QY	606	DLAVGF-GDKR-THQKVYVQGNIRYRPFIKKSVLR-----GYAKLGYG--NNLPFENFYAG	656				
DB	596	NAEIALPGSKLQYYATHNQITWFEPPLSKTFTMLGGEVGIAG-GYGRTKRPIPFENFYAG	654				
QY	657	GYGSRVGYDOSSILGPRSQAQVLTARRQOOTTLGEV--GGNATATFGSELILPLPFKGDWI	714				
DB	655	GLGSVRGYESGTLGPK-----VYDEYGEKISYGGNKKANVSAAELLFPMFGAKD-A	703				
QY	715	DOVRPVIFIEGGVDFDTGMDKOTIDLTPQKDPQATAEQNAKANRPLLTQDQLQYASAG	774				
DB	704	RTVRLSLFADAGSVWGGKIYDDNSSAT-----GGRVQNTIYGAGNTHKSTFTTNELRYASG	758				
QY	775	VGATWTPITGPLSISYAKPLUNKKONDQTDTPVQFQISVF	813				
DB	759	GAYTWLSPGLGPMKFSYAYPLKKKPEDEIQRFQOLGTTF	797				



**RESULT** 10  
D82000  
outer membrane protein OMP85 NMA0085 [imported] - Neisseria meningitidis (strain Z2491 s)  
C:Species: Neisseria meningitidis  
C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
C:Accession: D82000  
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel,  
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,  
Nature 404, 502-506, 2000  
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.  
A:Reference number: AB1775; MUID:2022556  
A:Accession: D82000  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-797 <PAR>  
A:Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; PIDN:CAB83401.1; PID:g737885  
A:Experimental source: serogroup A, strain Z2491  
C:Genetics:  
A:Gene: omp85; NMA0085  
C:Superfamily: protective surface antigen D-15

**Query Match** 26.4%; Score 1111; DB 2; Length 797;  
**Best Local Similarity** 32.4%; Pred. No. 2e-56;  
**Matches** 263; Conservative 155; Mismatches 361; Indels 38; Gaps 14;

Qy 10 QVSAMTMAVMVMSHTAAQAADFMANDITITGLQRVTIESLQSVLPPRLGQVYSENQLADG 69  
Db : : : l:l: : : ||| || : |||| : : : || : : : || : : :  
2 KLKQIASALAWLGISPLALADFTIQDIRVEGLQRTSTVFNYLPVKVGDTYNDTHGSAI 61  
Qy 70 VKALYTGNFSDVOYVHQEGRIIYQVTERPIAEINFEGRNLRIPEKGLEGLKNAGLAVG 129  
Db : : ||||| : |||| : : : ||| : ||| : : : ||| : : : ||| : : : |||  
62 IKSLYATGFDDVRVETADGOLLTVIERPTIGSLNTGAKMLQNDAIKKNLESFGLAQS 121  
Qy 130 QPLKAQTVMLETTLTWQYSQGYYNEIIVKQTMDCNRVKLDMTFAEGKPARYVDINI 189  
Db : : |||| : : : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||  
122 QYFNQATLNQAVAGLKEEYLGRGKLNIOTPKVTKLARNVRIDITIDECKSAKITDIEF 181  
Qy 190 IGNOHFSADALIDLVAIAIKDNKI-NPLSKADRYTOEKLVTSLENRAKYLNAGFVRPEIKD 248  
Db : : : ||| : : : | : : : : : : : : : : : : : : : : : : : : :  
182 EGNQVYSORKLMRQMSLTEGGITWLTFRSQNFQEQEAQMKEKVTDFYQNGGVDFRILD 241  
Qy 249 AKLNINEDKNRIFVEISLHGEEQYRFQOTQFLGNLT-YTQAELEALKFKAEFGSQAML 307  
Db : |||| : |||| : |||| : |||| : |||| : |||| : |||| : |||| :  
242 TDIOTNEDKTQTKITIVHEGGRFWGKVSEGTNEVPKAELEKLLTMKPGKWYERQQM 301  
Qy 308 EQTTNNISTFGDDGYXAOIRPVTRINDESRTVDVEYIDPVHPVYVRINTEGNPKTQ 367  
Db : : : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||  
302 TAVLGEITONRGSGSAGYASEISVQPLNAETKTVDFLHIEPGRKKIYVNEIHITGNKTR 361  
Qy 368 DEVLRRMRQLEGALASNQIKLSRARLMTGTFFKHVTVDTRPVPNSPDQDVNFVVEEQ 427  
Db : |||| : |||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||  
362 DEVVRELQMESAPYDTSKLRSKERVELLGYFDNQVQFADVPLAGTPDKVDLNMSLTER 421  
Qy 428 PSGSSTTAAGYSQGGVTFQFDVSYQNFMGTGKHVNASFSRSETREVISGLMTNPYFTVN 487  
Db : || : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||  
422 STGSLDLASAGVVQDTGLMVAGVSQDNLFCTGSAALRASRSKTTLNGSLSFDPYFTAD 481  
Qy 488 GVSQSLSCYLR--KTKYDNKNISNYVLDYSGGSLSYGYPDENQRISFGLNADNTKLHGG 545  
Db : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||  
482 GVS LGYDVYKAFDPRKASTSIKOYKTTTAGAGIRMSVPVTEYDRVNFGLVAEHLYTN-- 539  
Qy 546 RFMGISNVKOLMADGGKIQVDNNGIGPFKDHYTTYAILGMWNYSLLDRPVPFPPTQGMHSV 605  
Db : : : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||  
540 ---TYNKAPHYADFIIKYGKTDG-TDGSFEGWLKYKTVGWGRNKSTDALWPTNGYLTGV 595  
Qy 606 DLATVGF-GDK-THQKVYVQGNRIYRPFIKKS YLR-----GYAKLGYG--NNLPFYENFYAG 656  
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
596 NAEIALPGSKLQYYATHNTQWFPPLSKTTTLMLGEGVEGIAG-CYGRTEKPIEPFENPYGG 654  
Qy 657 GYGSVRGYDOSSLGPRSQAYLTARRGOQTTLGEVV--GGNALATFGSELILPLPFKGDWI 714  
Db : |||| : |||| : |||| : |||| : |||| : |||| : |||| : |||| : |||| : ||||  
655 GLGSVRGYESTLGPK-----VIDYBGEKISYGGNKKANVSAELFFPMPGAKD-A 703

[illegible]

Db 561 QSMK-----FKNGIITNDFDS-----FGWNINSLNRGYFPTKGVKASLGGGRVT 605  
QY 609 VFGDKTHQKVYQGNIRPF-----IKKSVLRGVAKLGYGN-NLPFFYENFYAGGVGS 660  
Db 606 IPGSDNKKYKLSADVQGFYPLDRHLWVVSASAKAGYAN-GFGNKRLPFYTYTAGGIGS 664  
QY 661 VRGYDQSSSLGRPSQAYLTARRGQOT---TLGEVVGGNALATFGSELILPLPFKGDWI-DQ 716  
Db 665 LRGFAYGSGIGNA-IYAEHNGNGCTFKKISSDVIGGNAITATASAEELIVPTPFVSDKQNT 723  
QY 717 VRPIETEGGQVDTT-GMDKQTDLTQFDPQATQATONAKAANRPLLTQDKQLRYAGV 775  
Db 724 VRTSLFVDAASVWNTKWKSDKSLD-----NNVLKSLPDYGGKSSRIASTGV 770  
QY 776 GATWYTPIGPLSISYAKPLNKKQNDQTDTVQFQIGSVF 813  
Db 771 GFQWQSPIGPLVFSYAKPIKKYENDVQFQFSGISGF 808  
RESULT 12  
JC4078  
protective surface antigen D-15 precursor - Haemophilus influenzae (type b)  
C:Species: Haemophilus influenzae  
A:Variety: type b  
C:Date: 30-Jun-1995 #sequence\_revision 14-Jul-1995 #text\_change 29-Sep-1999  
C:Accession: JC4078  
R:Flack, F.S.; Loomore, S.; Chong, P.; Thomas, W.R.  
Gene 156, 97-99, 1995  
A:Title: The sequencing of the 80-kDa D15 protective surface antigen of Haemophilus inf  
A:Reference number: JC4078; MUID:95255676  
A:Accession: JC4078  
A:Molecule type: DNA  
A:Residues: 1-797 <FLA>  
A:Cross-references: GB:U13961; NID:g537447; PIDN:AAA85645.1; PID:g537448  
A:Experimental source: type b  
A:Superfamily: protective surface antigen D-15  
C:Keywords: surface antigen  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-797/Product: protective surface antigen D-15 #status predicted <MAT>

Query Match 25.0%; Score 1052.5; DB 2; Length 797;  
Best Local Similarity 31.8%; Pred. No. 4.9e-53;  
Matches 260; Conservative 160; Mismatches 333; Indels 69; Gaps 20;  
QY 23 STHQAADFMANDITITGLQRTVIESLQSVLPFRLGQVSENOLADGVKALYATGNFSDV 82  
Db 14 TTTVFAAPEFAKDIRVDGVQDLEQQIRASLPVRAGQRTVDNDVANIVRSFLVSGRFDV 73  
QY 83 QVTHQEGRI-IYQVTERPLIAEINFEGRNLIPIKEGLQGLKNAGLAVGQPLKQATVOMIE 141  
Db 74 KA-HOEGDVLVSVVAKSIISDVKIKGNSVIPTEALKONLDANGFKVGDVLIREKLENEFA 132  
QY 142 TELTNQVISOGYKNTETVTKQTMLDGNRVLKDMTFAEGKPARVVDINIGNHFSADLI 201  
Db 133 KSKVHVASGRNATVEPIVNTLPNRAEILIQINEDDKAKLASUTFRKONESVSSSTLQ 192  
QY 202 DVLAIKDNKINPLSKADRYTQEKLVTSLENLRKAKYINAGVRFVEIKDAKININEDKNRIF 261  
Db 193 EQMELQDPSWWKL-WGNKPEGAQFEKDLQSIDRYLNNGYAKAQITKTDVQLNDEKTKVN 251  
QY 262 VELSLHEGEQYRGQTFQFLGNLYTQAELEALK-KFAEGRFSQAMLEQTNNTSTKFGD 320  
Db 252 VTIDVNEGLOYDLSARIIGNLGMASAEPLLSALHLNDFRSDIADVENAIAKALGE 311  
QY 321 DGYYYAQIRVTRINDESRRVDVEYYIDPVHVVYVRINFNTGNFKTQDQVLRERMQLEG 380  
Db 312 RYGTSATNSVPDFDANKTLATLVVDAGRRLTVKQLREFGTVVSADSTLRQEMRQEG 371  
QY 381 ALASNKIQLSRARLMRTGFFKFKHTVDTR--PVPNSPDQVDVNFVVEQPSGSSSTAAQY 438  
Db 372 TWYNSQVELGKRLDRTGFFE--TVENRIDPINGSNDEVVYVKERTGSGINFGICY 429

QY 439 SQSGGVTFQFDVSONNFMGTGKHVNASFSRSETREVYSLGMTNPYFTVNGYSQSLSGYR 498  
Db 430 GTESSISVQASVKQDNFLGTGAASVIAGTKNDYGTSVNLGYTEPYFFKDG--SLGNVNF 487  
QY 499 KTKYDNK---NISNVLDYSYSGSLSGYPIDENQIRISFGL-----NADNTKLHGGRPMGI 550  
Db 488 FENYDNSKSDTSSNYKRTTYGTSNVTGFPVNNNSYVGLGHTYNTKISNFALEYNRNLYI 547  
QY 551 SNVQQLMADGGKIQVDNNGIPDPFKHDYTYNAILGWYSSLDPRVFPPTQGHSHVD--LT 608  
Db 548 QSMK-----FKNGIKTNDPDFS-----FGWYNLSNRGYFPTKGVKASLGGGRVT 592  
QY 609 VFGDKTHQKVYQGNIRPF-----IKKSVLRGVAKLGYGN-NLPFFYENFYAGGVGS 660  
Db 593 IPGSDNKKYKLSADVQGFYPLDRHLWVVSASAKAGYAN-GFGNKRLPFYTYTAGGIGS 651  
QY 661 VRGYDQSSSLGRPSQAYLTARRGQOTTLG-----EVVGGNALATFGSELILPLPFKGDW 713  
Db 652 LRGFAYGSGIGNA---IYAEYNGSGTGTFFKISSDVIGGNAIATASAEELIVPTPFVSDK 708  
QY 714 I-DOVRPVIETEGGQVDTT-GMDKQTDLTQFDPQATQATONAKAANRPLLTQDKQLRY 771  
Db 709 SQNTVRTSLFVDAASVWNTKWKSDKNGLESVDLK-----RLPDYGGKSSRIRA 755  
QY 772 SAGVGATWYTPIGPLSISYAKPLNKKQNDQTDTVQFQIGSVF 813  
Db 756 STGVGFQWQSPIGPLVFSYAKPIKKYENDVQFQFSGISGF 797  
RESULT 13  
D71726  
outer membrane protein omp1 (omp1) Rpl160 - Rickettsia prowazekii  
C:Species: Rickettsia prowazekii  
C:Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 03-Nov-2000  
C:Accession: D71726  
R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark  
Nature 396, 133-140, 1998  
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.  
A:Reference number: A71630; MUID:99039499  
A:Accession: D71726  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-768 <AND>  
A:Cross-references: GB:AJ235270; GB:AJ235269; NID:g3860572; PIDN:CAA14627.1; PID:g386  
A:Experimental source: strain Madrid E  
C:Genetics:  
A:Gene: omp1; Rpl160  
C:Superfamily: protective surface antigen D-15  
Query Match 17.9%; Score 754; DB 2; Length 768;  
Best Local Similarity 26.3%; Pred. No. 8.4e-36;  
Matches 216; Conservative 145; Mismatches 383; Indels 76; Gaps 20;  
QY 7 KGFOVSAMTMAVMVMSTHAQAADFMANDITITGLQRTVIESLQSVLPFRLGQVSVENQL 66  
Db 2 KIISIKSLTILLTIFYHISFADYVIRKTIIEGNHVRSTIESYLKLVNGETYNNKSKE 61  
QY 67 ADGVKALYATGNSDQVY-HQEGRIIYQVTERPLIAEINFEGRNLIPIKEGLQGLKNAG 125  
Db 62 DEAIKRLIATSLERNIMYITNDGNLIVNVTETPTFSSVVFSGNSKTKTNILAKEIYTMS 121  
QY 126 LAVGQPLKQATVOMIETELTNQVISOGYKNTETVTKQTMLDGNRVLKDMTFAEGKPARV 185  
Db 122 ---GESLSQAKIELDKVKKILEIYKSRGRFSTKTPPKLSLENNRVKIVFDIAEGPKTVIK 178  
QY 186 DINIGNQHSFDADLDVLAIKONK-INPLSKADRYTQEKLVTSLENLRKAKYINAGVRF 244  
Db 179 SIYFSNGEYSDSELKSIVLTKEISWFRFLESNDYDPRVEYDKELLREFYQSVGPADE 238  
QY 245 EIKDAKININEDKNRIFVELSLHEGEQYRGQTFQFLGNLYTQ-AELEALLKFAEGRFS 303

```
Db 239 RVISASVALNDKTEYFTTYSIEEGEKYRFGNVTIDNKLNTINIKOLKNIVNIKOGKIFN 298
Qy 304 QAMLEQTTNNISTKFGDDGYIAQRPVTRINDESRITVDVEYIDPVHPVYVRRINFTGN 363
Db 299 MKTVDDIAEKICEYFANGYPAVNYPIKND-NHTADIFIEKADKVINKINIINN 357
Qy 364 FKTODEVLRMRQLEGALASNQKLTOLSRARLMRTGFFKHVTVDRPVPNSDQVDVNFV 423
Db 358 LKTEDHVIRRAFKIBEGDVMRSYTEKGERNLRLNDYFEKVSISLAQT-KAKDKYDVNVE 416
Qy 424 VEEOPSGSSTIAAGYSQSGGVTFQDVSONNFMGTGKHVNASFSETRREYVSLGWTNPY 483
Db 417 VDEKSTSIGDLGNTAGGLFGRFSFLERNLVGTGKLLNAGVQVSKNSTSYGCITDHP 476
Qy 484 FTVNGVSQSLSGY----RKTNDKNKNSINYVLDYSGSLSGYGPIDE--NORISFGUNA 537
Db 477 FLDRDLSLVNAFRNYTRGASVLTNTDQSYKLSIGYKISLGYDMKEDLSHEIDYLIKR 536
Qy 538 DNTKLHGRFRGINSVVKOLMADGGKIQVDNNGIPDFKHDTYTYNAILGNYSLSLDRPVP 597
Db 537 DILSAPSP-----SNSIFLNEQMGKLTISAIG-----HTIT-----YDQTDNKKVP 577
Qy 598 TQG--MHSVDLTGVFGDKTHQKVYVYQGNIRYRPFITKKSVLRCYAKLG-----YGNLPLF 649
Db 578 KNGILVSTQTEFAGVGGDNKKYIKHEIECKFYKSFNNKVTUKLSAAGGDMAGLGKMWRI 637
Qy 650 YENFYAGGYGVRGYSQSSLGPRSQAYLTARRGQOTTILGEVVGYNALATFGSELILPLPF 709
Db 638 SDRFNIGDY-SLRFPASGVGPR-----EKNTEGLGGERYYTFFSTELNFPVP 685
Qy 710 KGDWIDQVRPVIFIEGGQVFTTGMKQOTIDLTQPKDPOATAEQNAKANRPLLTQDKOL 769
Db 686 PEEF--NFTGAVFIDLGSVMG-VGLNKK-----QYKTPNG-----FYNDQSL 724
Qy 770 RYSAGVGATWTPTIGPLSISYAKPLNKKQNDQTDVQFOI 809
Db 725 RASVGFGIWTRFAPIRMDMGFPFKKQYD--DFQNEHL 762

RESULT 14
D97527
ompl protein precursor (U51683) [imported] - Agrobacterium tumefaciens (strain C58, Cere
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C:Accession: D97527
A.: Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Ourullo, B.; Goldman,
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; PMID:11743194
A:Accession: D97527
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-774 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK87173.1; PID:g15156447; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_2554
A:Map position: circular chromosome

Query Match 17.5%; Score 734; DB 2; Length 774;
Best Local Similarity 26.4%; Pred. No. 1.2e-34;
Matches 219; Conservative 150; Mismatches 354; Indels 106; Gaps 22;

Qy 15 TMAYMVMTSHAQAADFANADITITGLQRTVIESLQSVLPRLGOVSVSENLADQVKALY 74
Db 22 SVAGLGVLASAGVANAAVISIDVRGASGADSVRSNITTAPGNFNSDIDESVKRLY 81
Qy 75 ATGNFSDQVYHQEGRIIVQYTERPLIAEINFEGNRLIPKQGLQBLKAGLAVGQPL-- 132
Db 82 ATGFYSNVMRSVSGSLVTWTVNENQLVNQVFNKRIKDD-----KLAVIQVQTPMGP 135
Qy 133 -KQATVQMIEFTLNTQYISQGYNTETITVKQTMLDGNRVKLDMTFAEGKPARVVDINIIG 191
```

```
Db 136 FNOAVTADIAIRKEAYSAGRSDEVTITQTVSVGGRVNIAPVINEGERTKIGRIDFIG 195
Qy 192 NOHFSADALIDLVL-AIKDNKINPLSKADRYTQEKLVTSLENLRAKYLNAGVRFPEKDAK 250
Db 196 NNSYSDGRLAAYINTKKSNNLSFLTRKDVYNEDKLRADEEALRQFYNNRGYADFVYSSD 255
Qy 251 LMINEDKNRIFVEISLHEGEQYRFGQTQFLGNLTYYQ-AELEALLKFAEEGFSQAMLEQ 309
Db 256 AVLDESKENEYTISITVDEGKKYDFGNVAVESTVPGVDSGSELOGLVETROGASYSAREVOQ 315
Qy 310 TTNNISTKFGDDGYIAQRPVTRINDESRITVDVEYIDPVHPVYVRRINFTGNFTQDE 369
Db 316 SMEAISKRVRAGEYFPFARYTPRGDRDMSGNTIGVTYIVDQGERAYVERIEIRGNTTRDY 375
Qy 370 VURRMRQLEGALASNOKLTOLSRARLMRTGFFKHVTVDRPVPNSDQVDVNVVVEEQPS 429
Db 376 VIRREFDISEGDAFNQTTITAAKRLLEALGYFSKVNISTAG-GSAPDRVVIVVVDVEDQST 434
Qy 430 GSSTIAAGYSQSGGVTFQDVSONNFMGTGKHVN-ASFRRSETREYVSLGWTNPYFTVNG 488
Db 435 GSFGIGAGYSQNDGVLLEASVEEKNFLGRQYIRVAAGAGEDDARTYLSLFTPEYF---- 490
Qy 489 VSQSLSGYRKYTKYDNKNISNVLDSYSGSLSGYGPIDENQRISFGNLADNATKLHGGRFM 548
Db 491 LGYRLAAGFDLEFKNOSKSEDIYNYDEQGFALRVTAITENLSITPKY----- 537
Qy 549 GISNVKQLMADGGKIQVDNNGIPDFKHDTYTYNAIL--GW-----NYSSLDRPVF 596
Db 538 ----TYKQINVEGKGWONNANLAE-----PYOALIRGEDWTQSILSNLTNLYNTDNRNM 588
Qy 597 PTQGMHSHSVDLNVGF----GDKTHQKVYVYQGNIRYRPFIKK----SVLRGYA--KLGYGNN 646
Db 589 PREGVQAA--LTNEFAGLGSDSEYKIKYAKARYYYTSLDSDEYDVIGSLTQAGHVMFTGDN 646
Qy 647 LPFENFYAGGYGVRGYSQSSLGPRSQAYLTARRGQOTTILG-EVVGGNALTATFGSELIL 705
Db 647 LLVDFQFKFGG-RQVRGFKNDGIGPR-----IGSDSIGGTTYYFAASAEEVTA 691
Qy 706 PUP-FKGWDIDQVRPVIFIEGGQVFTTGMKQOTIDLTQPKDPOATAEQNAKANRPLLT 764
Db 692 PMPGVPEDE--GLRLAGFVDACTMGKNYVSTQTV-----KD----- 726
Qy 765 QDKQLRYSAGVGATWTPTIGPLSISYAKPLNKKQNDQTDVQFOIGSVF 813
Db 727 -DNSIRASAGIGVMASPPGPIRVDYAIPAKEDYDEEQRFREGMSNTF 774

RESULT 15
AE2746
group 1 outer membrane protein precursor ompl [imported] - Agrobacterium tumefaciens
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AE2746
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCl
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Fam
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AE2746
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-774 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAL42387.1; PID:g17739796; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: ompl
A:Map position: circular chromosome
```

```
Query Match      17.5%; Score 734; DB 2; Length 774;
Best Local Similarity 26.4%; Pred. No. 1.2e-34;
Matches 219; Conservative 150; Mismatches 354; Indels 106; Caps 22;

QY 15 TMAYMVMSTHAQAAMANDITITGLQRTVIESLQSVLPRLGQVVSLENOLADGVKALY 74
Db 22 SVAGLVLASAGVANAIVISKIDYRGASGADSVRNITITAPGNFNSDIDESVRLY 81

QY 75 ATGNFSDVQVYHQGRITTYQTERPLIAEINFEGRNLIPKEGLQGLKNAGLAVGQPL-- 132
Db 82 ATGYFSNVSMRVSGSTLVVTVNENQLNVQVFNKRIKDD-----KLAGIVQTQPMGP 135

QY 133 -KOATVQMIETELTNQYISQGYNYFEITVKTMLDGNRKLDMTFAGKPARVVDINIIG 191
Db 136 FNOAIVTADTARIKEAYSAIGRSDVEITTTQVSGOGRVNIATVINEGERTKIGRIDFIG 195

QY 192 NQHFSDADLIDLVL-AIKDKINPLSKADRYTQEKLVTSLENLRKAKYLNAGFVRFKIDAK 250
Db 196 NNSYDGLAALAVINTKKSMLSFTRKDVYNEDKLRADEALROFYNNRGADFRVVSDD 255

QY 251 LNINEDKNRIFVEISLHEGBOYRFGQTFGLNLTYYTO-AELEALLKPKAEFGFSQAMLEQ 309
Db 256 AVLDESKNEXTISTITVDEGKKYDFGNVAVESTVPGVDGSELQGLVETRGQASYSAKEVQQ 315

QY 310 TTNNISTKFGDGYVYIAQIRPVTRINDESRIVDEYVIDPVHPVYVRRINFTGNFKTQDE 369
Db 316 SMEAISRVVAGEGYPFARVTPRGDRMSGNTIGVTYIVDQGERAYVERIERGNTRTRDY 375

QY 370 VLRRMRQLLEGALASNOKIQLSRARLMRTGFFKHVTVDTRPVNPSPDQVDVNFVVEOPS 429
Db 376 VIRREFDISEGDAFNQPIITAARKRLEALGYFSKVNISTAG-GSAPDRVIVVDVEDQST 434

QY 430 GSSTIAAGYSQSGGVTFQFDVQNNFMGTGKHVN-ASFRRSETREYVSLGTMNPYFTVNG 488
Db 435 GSGFAGAGYSQNDGVLEASVEEKNFLGRGYIRVAAGAGEDDARTYSLSFTEPYF---- 490

QY 489 VSQSLGYYRKTVDNKNISNVYVLDVSGSLSYGPIDENQRIQSFGLNADNTKLHGGRFM 548
Db 491 LGYRLAAGFDLKNQSKSEDIYNYDEQGFALRVTPITENLSTTFKY----- 537

QY 549 GISNVKOLMADGGKIQVDNNGIPDFKHDTTYNAIL---GW-----NYSLSLRPVF 596
Db 538 ---TYKQINYEKGKDWQNNANLAE-----PYQALIRGEDWQTOSILSNTLNYNTLDDRN 588

QY 597 PTQGMHSVDLTGVF----CDKTHQKVYQGNIRYRPIKX-----SVLRGYA--KLGYGNN 646
Db 589 PREGWQAA--LTNEFAGLGDSEYKIIYAKARYYYTSLSEYDVIQSLTGQAGHVMPTGDN 646

QY 647 LPFYENFYAGCYGSRGYDQSSILGPRSQAYLTARRGQQTTLG-EVYGGNALATFGSELIL 705
Db 647 LLVFDQFKFGG-QVRGFRKNDGIGPR-----IGSDSIGGTYTYFAASAETA 691

QY 706 PLP-FKGDWTDQVRPVIIEGGQVFTTGMKQTDIDTQFKDPQATAEQNAKANRPLLT 764
Db 692 PMPGVPEDE--GLRLAGFVDAGMTYGNKVSTSTV-----KD----- 726

QY 765 QDKQLRYAGVATWYTPIGPLSISYAKPLNKKQNDQTDVTFQIGSVF 813
Db 727 -DNSIRASAGIGVWASPFGRVIRDYAIPAKEDYDEQRFREFGNGNTE 774
```

Search completed: September 5, 2002, 09:56:02  
Job time: 281 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 10:03:36 ; Search time 36.65 Seconds  
(without alignments)  
858,908 Million cell updates/sec

Title: US-09-701-711-4  
Perfect score: 4202  
Sequence: 1 MRNSYKFGQVSAMTAVMM.....LNKKQNDQTDVQFIGSVF 813

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query %	Match	Length	DB ID	Description
1	1210	28.8	810	1	UP05_ECOLI	P39170 escherichia
2	1053.5	25.1	795	1	D152_HAEIN	P44935 haemophilus
3	1052.5	25.0	793	1	D153_HAEIN	O32629 haemophilus
4	1052.5	25.0	797	1	D151_HAEIN	P46024 haemophilus
5	641	15.3	617	1	Y237_BUCAI	P57331 buchnera ap
6	256.5	6.1	578	1	YTFM_HAEIN	P44038 haemophilus
7	227	5.4	577	1	YTFM_ECOLI	P39320 escherichia
8	197.5	4.7	469	1	C651_HUMAN	Q9Y512 homo sapien
9	171	4.1	463	1	U140_DROME	Q9V784 drosophila
10	157.5	3.7	1577	1	HLYA_PROMI	P16466 proteus mir
11	154.5	3.7	1256	1	MRP_STRSU	P32653 streptococc
12	149	3.5	1902	1	P3P_LACLC	P15292 lactococcus
13	148	3.5	475	1	YDF6_SCHPO	Q10478 schizosacch
14	146	3.5	826	1	YEBB_ECOLI	P33341 escherichia
15	146	3.5	1902	1	P1P_LACLC	P16271 lactococcus
16	143	3.4	1902	1	P2P_LACPA	Q02470 lactobacill
17	142.5	3.4	705	1	PPCF_FLAME	P27195 flavobacter
18	140.5	3.3	1902	1	P2P_LACLC	P15293 lactococcus
19	140	3.3	2201	1	TENA_HUMAN	P24821 homo sapien
20	139.5	3.3	1324	1	CUT3_SCHPO	P41004 schizosacch
21	137.5	3.3	735	1	FCT_ERWCH	Q47162 erwinia chr
22	136	3.2	917	1	HXA3_HAEIN	P45355 haemophilus
23	135.5	3.2	1332	1	XKDO_BACSU	P54334 bacillus su
24	134.5	3.2	866	1	YCBS_ECOLI	P75857 escherichia
25	134.5	3.2	1178	1	CIAC_BACTK	P03068 bacillus th
26	134	3.2	655	1	YKDA_MYCCA	P45615 mycoplasma
27	133.5	3.2	937	1	C532_ECOLI	P15484 escherichia
28	133	3.2	398	1	YLB1_CAEEL	P46576 caenorhabdi
29	133	3.2	1048	1	AGOL_ARATH	O04379 arabidopsis
30	132.5	3.2	863	1	Y8TO_ECOLI	P33924 escherichia
31	130.5	3.1	1957	1	YD86_SCHPO	Q10411 schizosacch
32	128.5	3.1	905	1	HXAL_HAEIN	P44602 haemophilus
33	127.5	3.0	839	1	YDDB_HAEIN	P45182 haemophilus

#### ALIGNMENTS

RESULT 1  
UP05\_ECOLI

ID UP05\_ECOLI STANDARD; PRT; 810 AA.  
AC P39170: P39181; P77465;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DE Unknown protein from 2D-page spots M62/M63/O3/O9/T35 precursor.  
GN YAEI OR B0177 OR Z0188 OR ECS0179.  
OS Escherichia coli, and  
OS Escherichia coli O157:H7.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562, 83334;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Sequence 277:1453-1474(1997).

[2]  
RP SEQUENCE FROM N.A.  
RC Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,  
RA Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,  
RA Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,  
RA Davis R.W.;  
RP Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
[3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / ATCC 700927;  
RX MEDLINE=21074935; PubMed=11206551;  
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamowski K.,  
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
RA Welch R.A., Blattner F.R.;  
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";  
RL Nature 409:529-533(2001).  
[4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / RIMD 0509952;  
RX MEDLINE=21156231; PubMed=11258796;  
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
O157:H7 and genomic comparison with a laboratory strain K-12.";  
RL DNA Res. 8:111-22(2001).  
[5]  
RP SEQUENCE OF 21-32 AND 351-362.

34 127 3.0 484 1 YNC6\_YEAST  
35 127 3.0 922 1 DPO1\_RICFE  
36 126.5 3.0 1258 1 ICEN\_ERWHE  
37 126 3.0 925 1 VPH\_BHPH1  
38 126 3.0 1034 1 ICEN\_PANAN  
39 125.5 3.0 880 1 LYTD\_BACSU  
40 125.5 3.0 2022 1 ANTI\_ONCVO  
41 124.5 3.0 1341 1 VG37\_BPT2  
42 124.5 3.0 1754 1 PMPB\_CHLTR  
43 124.5 3.0 2366 1 TOXB\_CLODI  
44 124 3.0 1630 1 MSP1\_PLAFK  
45 124 3.0 1639 1 MSP1\_PLAFW

P53969 saccharomyc  
Q9raa9 rickettsia  
P16239 erwinia her  
P51735 bacterioph  
Q47879 pantoea ana  
P39848 bacillus su  
P21249 onchocerca  
P07067 bacterioph  
O84418 chlamydia t  
P18177 clostridium  
P04932 plasmodium  
P04933 plasmodium

```
RC STRAIN-K12 / BMG2;
RX MEDLINE=97443975; PubMed=9298646;
RA Link A.J., Robison K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
RL in the genome of Escherichia coli K-12";
RC Electrophoresis 18:1259-1313(1997).
CC -!- SUBCELLULAR LOCATION: Outer membrane (Potential).
CC -!- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; AE000127; AAC73288.1; -
CC DR EMBL; U70214; AAB06606.1; -
CC DR EMBL; AE005193; AAG54479.1; -
CC DR EMBL; AP002550; BAB33602.1; -
CC DR SWISS-2DPAGE; P39170; COLI.
CC DR EcoGene; EGI2676; Yaet.
CC DR InterPro; IPR000184; Bac_surfAg_D15.
CC Pfam; PF01103; Bac_surface_Ag; 1.
CC KW Outer membrane; Signal; Complete proteome.
CC FT SIGNAL 1 20
CC FT CHAIN 21 810
CC FT SEQUENCE 810 AA; 90552 MW; DDC64C6D341664EB CRC64;
CC -----
CC Query Match 28.8%; Score 1210; DB 1; Length 810;
CC Best Local Similarity 34.0%; Pred. No. 1.1e-63;
CC Matches 287; Conservative 149; Mismatches 330; Indels 78; Gaps 15;
CC -----
QY 14 MTMAVMMVMS-----THQAADPMANDITITGLQRVTIESLSQVLPFRLCQVVSQNOLA 67
DB 1 MAMKLLIASLFFSATYVGAEGFVKDIHFEGLRQVAVGAALLSMPRTGDTVNDDEIS 60
QY 68 DGVALYATGNFSDQVYHQBGRYIYQVTERPLIAEFNIGRNLIPKEGLQELKNAGLA 127
DB 61 NTRALFATGNFEDVRVLRDGDTLLQVQKERTTIASITFGSGKSVKDDMLKQNLQESVGR 120
QY 128 VGQPLKQATQVMIETELTNQYISOGYNTETVTKQTMLDGNRVKLDMTFAEGKPARVVDI 187
DB 121 VGESLDRRTIADIEKGLDEFYISVGYKSASVKAIVTLPRLNRVLDKLVFQEGVSAEIQOI 180
QY 188 NIIGNQHFSDADLDVLAIKD-----NKINPLSKADR-YTQEKLVTSLENLRAKYLNAGF 241
DB 181 NIVGNHFTTDELISHFQLRDEVPWNVV-----GDRKYQKQKLAGDLETLRSYLDRGY 235
QY 242 VRFKIDAKLINDEKRIKRIEISLHGEQYRFGQTQFLGNLTVTQAEALALKFKAEEG 301
DB 236 ARFNIDSTQVSLTPDKKGIYVTNITEGDQYKLSGVEVSGNLGHSAEIEQLTKIEPGL 295
QY 302 FSOAMLEOTTNNISTKFGDGYIAOIRPYTRINDESRDVTVEYIDPVHPVYVRNFT 361
DB 296 YNGTKVTKMEDDKLLGRGYAYPRVQSMPEINDAKTVKLRVNDVAGNRFYVRKIRFE 355
QY 362 GNFKTQDEVLRREMRQLEGALASNQIKLSRLMRRTGFFKHVTVDRPVPNSPDQVDVN 421
DB 356 GNDTSKDALVRREMRQEGAWLGSGLVDQKGERLNRGLFFETDQTRVPGSPDQVDV 415
QY 422 FVVEEQPSGSSSTAAGYSQSGVTFQFDVSONNFMGTGKHNASFSRSRETVYSLGTMN 481
DB 416 YKVKERTNIGFNFGYIGTGESGSFQAGVQDDNWLGTGAVGNGTKNDYQTYAELSVTN 475
QY 482 PYFTVNGVSQSLSGYRKTKYDNKNISNYVLDSVGSLSYGYPTDENQRIISFGLNADNTK 541
DB 476 PYFTVDGSLGGRLFYNDFFADADLSDYNTKSYGTDVTLGFFINEVNSURAGLGVHNS 535
QY 542 LHGGRFMGISNVKQLMA-----DGGKIQVDNNGIPDFKHDTYTYNAILGNWYSLSLRPV 595
```

```
Db 536 L-----SNMQPOVAMRWYLSAGEHPSTSDQDSFKTDDFTFN--YGWTYNKLDRCY 585
QY 596 PPTQG--MSHSDVLDIVGFGDKTHQKVYOGNIYRPF--IKKSVLRGYAKLGYGNL---- 647
Db 586 PPTDGSRYNLTKGVTPGSDNEYRYKVTLDATATYVPTDDHKKWVLGRTRWGYDGLGGKE 645
QY 648 -PYENFYAGGVSGVSGYDQSSLSGPRS-----QA-----YLTARRCOQTTL---GEV 690
Db 646 MPFYENFYAGGSSTVRGFSQNTIGPKAVYFPHOASNYDPDYDECATQDGAOKDLCKSDDA 705
QY 691 VGGNALATFGSELILPLPFKGD-WIDQVRPVIFIEGGQVFDFTGMDKQTDILDTQFKDQQA 749
Db 706 VGGNAMAVASLEFPTPTFISDKYANSVRTSEFFDMGIVWDINWDSQSGYPYSDP--- 763
QY 750 TAEQNAKAAANRLLTQDKQLRYSAGVGATWYTPIOPLSISYAKPLNKKQNDQTDVQFOI 809
Db 764 -----SNIRMSAGIALQWMSPLGLPVFSYAQPFPKYDGDGAEOQFNI 806
QY 810 GSVF 813
Db 807 GKTW 810
RESULT 2
ID D152.HAEIN STANDARD; PRT; 795 AA.
AC P44935;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-Ag) (Outer membrane protein D15).
GN HI0917.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kienle C.A., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; U32773; AAC22575.1; -
CC DR TIGR; HI0917; -
CC DR InterPro; IPR000184; Bac_surfAg_D15.
CC Pfam; PF01103; Bac_surface_Ag; 1.
CC KW Antigen; Outer membrane; Signal; Complete proteome.
CC FT SIGNAL 1 19
CC FT CHAIN 20 795
CC FT SEQUENCE 795 AA; 87478 MW; B85691FC22E6D44 CRC64;
```

Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
Haemophilus.  
NCBI\_TaxID=727;  
[1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=PAK 12085;  
RX MEDLINE=97427952; PubMed=9284140;  
RA Loosmore S.M., Yang Y.P., Coleman D.C., Shortreed J.M., England D.M.,  
R Klein M.H.;  
RT "Outer membrane protein D15 is conserved among Haemophilus influenzae  
species and may represent a universal protective antigen against  
invasive disease.";  
RL Infect. Immun. 65:3701-3707(1997).  
CC -1- SUBCELLULAR LOCATION: Outer membrane.  
CC -1- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation  
at the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC -----  
DR EMBL; U60834; AAB61977.1; -  
DR InterPro; IPR000184; Bac\_surfAg\_D15.  
DR Pfam; PF01103; Bac\_surface\_Ag; 1.  
KW Antigen; Outer membrane; Signal.  
FT SIGNAL 1 19 POTENTIAL.  
FT CHAIN 20 793 PROTECTIVE SURFACE ANTIGEN D15.  
SQ SEQUENCE 793 AA; 87511 MW; 51BFBDB2036801A14 CRC64;  
  
Query Match 25.0%; Score 1052.5; DB 1; Length 793;  
Best Local Similarity 31.98; Pred. No. 1.9e-54;  
Matches 260; Conservative 160; Mismatches 333; Indels 65; Gaps 21;  
  
QY 23 STHAAQADFMANDTITGLQRTVIESQLSPVFLRGQVSVENOLADGVKALYATGNSDF 82  
DB : || || || || : || : || : || || || || || : || : || : || ||  
14 TTTVEAAPFVAKDIRVDGVQGLDEQQIRASLPRVGRQVTDNDVANIVRSILFVSGREDDV 73  
QY 83 QVYHOEGRI-IYQVTERPLIAETNFEGNRLPIKPLEGLKQGLKNAGLAVGQPLKQATVMIE 141  
DB : || || || : || : || : || : || : || : || : || : || : || : ||  
74 KA-HQEGDLVSVVAKSIISDKVKGISVPTALQNLDANGFKVGVDLIREKLNEFA 132  
QY 142 TELTNQVISQGYNTEITVKTMLDGNRVKLDMTFAEKKPARVVDINIGNOHESDADLI 201  
DB : || || || || : || || : || : || : || : || : || : || : || : ||  
133 KSVKEHYASVGRYNATVEPIVNTLPNNRAEILIQINEDDKAKLASLTFKGNESVSSLTQ 192  
QY 202 DLVAIKDNKINPLSKADRYTQEKLVTSLENLRAKYNAAGFVRFEIKAKLININEDKNRIF 261  
DB : : : : : : : : : || || || || || : || || : || : || : || : ||  
193 EQMELQPDSSWKL-WGNKFECAQEKDLQALRDYLLNNGYAKAQITKTDVQLNDEKTKVN 251  
QY 262 VEISLHGEQYRFQGTQFLGNLTVTQAEALALK-FKAEGBFSAOMLEQTTNNITSTKFGD 320  
DB : || || || : || || || || || || || || || || || || || || : || : ||  
252 VTIDVNEGLQDLRSARIIGNLGMSEAPELLSALHLNDFRRSDIADVENAIKALGE 311  
QY 321 DGYYYAQRVTRINDESRVDVEYYIDPVHPVYVRINFTGNPKTODEVLRREMROLEG 380  
DB || : : : : : : : : : || : || : || : || : || : || || || || ||  
312 RGVGNTTVNSVPDPDDANKTLAITFVVDAGRRLTVRLRFEQNTVSADSLRQEMRQEG 371  
QY 381 ALASNQKIQLSRARMLRTGFKKHVTVDR--PVPNSPDQVDNVFVBEQPSGSSTIAAGY 438  
DB : || : || : || || || || || : || : || || : || : || : || : || ||  
372 TWNSQLVELGKILDRTFGEF--TVENRRDIPNGSNDVDVYKVKERTGTSINFGIGY 429  
QY 439 SQGGGVTFQFDVSNQPMGTKHKHNASFSRSETREYVSLGMTNPFTVNGVSQSLSYR 498  
DB : || : || : || || || : || : || : || : || || || || || || || ||  
430 GTESGISQYTSIKQDNLFLGTGAASVIAGTKNDYGTSVNLGYTEPYFTKDGV--SLGNGIF 487  
QY 499 KTKYDNK---NISNYLDSTYGGSLSYGYPIDENQRIISFGL-----NADNTKLHGGRPMGI 550  
DB || || || : || : || : || : || : || : || : || : || : || : ||  
488 FENYDNSKSDTSSTNYKRTTYGSNVTFLGFPYNNNSYVGLGHTYNTKINSFALEYNNRLYI 547

Query Match	25.1%	Score 1053.5	DB 1	Length 795
Best Local Similarity	31.8%	Pred. No. 1.6e-54		
Matches	260	Conservative 159	Mismatches 337	Indels 63
	Gaps			
QY	23	STHAQAADFMANDITITGLQRTVTIESLQSVLPFRGLQGVVSENQLADGVKALYATGNFSDV	82	
DB	14	TTTFAAPFAVKDIRVDGVQGDLEQQIRASLPRVAGQRTVDNDVANIYRSLEFVGREDVV	73	
QY	83	QVTHQEGRI-IYQVTERPLAEINFEGRNLIPKGLQEGKLNAGLAVGQPLQATVQMI	141	
DB	74	KA-HQEGDLVSVVAKSIISDVKIGNSIIPTTALQNQLDANGKVGVDVLIRKLENEFA	132	
QY	142	TELTNOYISOGYNTETITVQTMIDGNRVKIDMTFAEKPAPVVDIINIINQHFSDADLI	201	
DB	133	KSVKEHVASGRYNAVTEPIVNTLPNNRAELIQLINEDDRAKLASLTKFGNESVSSLTQ	192	
QY	202	DVLAIKDNKINPLSKADRYTQEKLVTSLENLKAYLNAGFVRFEEKAKLNINEDKNRIF	261	
DB	193	EQMELQPDSSWMKL-WGNKFEQAQPEKDLQSDIRYLYNNGYAKAQTITKDVOLNDEKTKVN	251	
QY	262	VEJSLHEGEYRFQQTQFLGNLTVQAELEALKL-FAEEGFSQAMLEQTTNNSTTFEGD	320	
DB	252	VTIDVNEGLODLSARIIGNLGMSEALEPLLSALHLNDFRRSDADVENAKALIGE	311	
QY	321	DGYIAQIRPVTRINDESRTVDVEYYIDPVHPVVRINFTGNFKTODEVLRREMRQLEG	380	
DB	312	RGYGSATVNSVPDDANKTLAITLVVDAGRRLTVQLRFGNIVSADSTLRQEMRQEG	371	
QY	381	ALASNQIKLSRAKLMRTGFKKHVTVDR--PVPNSPDQVDVNFVVEEQPGSGSTIAAGY	438	
DB	372	TYNSQLVELGKIRLDRTGFFE--TVENRRIDPINGSNDEVDVYKVKERNGTSGINFGICY	429	
QY	439	SQSGGVTFQFDVSONNFMGTGKHVNASFSRSETREYVSLGMTNPTFTVNGVSQSLGYR	498	
DB	430	GTESGISYQASVKQDNFLGLGAAVSIAGTKNDYGTSVNLGYTEPYFTKDGV--SLGGNVF	487	
QY	499	KTYDNK--NISNYLVDYSYGSLSYGPIDENORISFGL-----NADNTKLHGGRFMGI	550	
DB	488	FENVDSKDTSSNYKRTTYGSNVTLGFPVNNNSYVGLGHTYKINSNFALEYNRNLYI	547	
QY	551	SNVKQLMADGGKIQVDNNGIPDFKHDTYTNAILGNWYSSLDPRVFPFQGMHSVD--LT	608	
DB	548	QSMK-----FKNGIKTNDPDFS-----FGWNYNSLNRGYFTPTKGVKASIGLRVT	592	
QY	609	VGFGDKTHQKVVYOGNYIRPF-----IKKSVLRGYAKLGYGN-NLPFFYENFAGGYGS	660	
DB	593	IPGSDNKYYKLSADVQGYFLDRDHLWWSAKASAGYAN-GFGNKRLPFTYOTYTAGGIGS	651	
QY	661	VRGYDOSSLGRSQAYLTARRGQOT---TLGEVVGGNALATFGSELILPLPFKGDWT-DQ	716	
DB	652	LRGFAYGSIGPNA-IYAEHGNCGTFFKISSDVIGGNAITASRELIVPTFPFVSDKQONT	710	
QY	717	VRPVIFIEGGQVDTT-GMDKQITIDLTFQDPQATAEQNAKAANRPLLTDQKURYSAGV	775	
DB	711	VRTSLFVDAASVWNTKWKDSKGLD-----NNVLKSLPDYGKSSRIRASTGV	757	
QY	776	GATWYTPIGPLSLTSYAKPLNKKQNDQTDTVQFQIGSVF	813	
DB	758	GFWQSPIGPLVFSYAKPIKKYENDDDVEQFQFSIGGSF	795	

RESULT 3

D153\_HAEIN

ID

D153\_HAEIN

STANDARD;

PRT;

793 AA.

AC

O32629;

DT

16-OCT-2001 (Rel. 40, Created)

DT

16-OCT-2001 (Rel. 40, Last sequence update)

DT

16-OCT-2001 (Rel. 40, Last annotation update)

DE

Protective surface antigen D15 precursor (80 kDa D15 antigen)

DE

(D-15-Ag) (outer membrane protein D15).

OS

Haemophilus influenzae.



```
QY 551 SNVKQLMADGGKIQVDNNGIPDFKHDDYTYNAILGNVSSLDPRVFPFQGMHSDV--LT 608
Db 548 QSMK-----FKNGIKTNDDEFD-----FCWNYNSLNRGYFPTKGVKASLGGRTV 592
QY 609 VGFEGDKTHOKVV--YQG-----NIYRPFIKKSVLRGVAKLGYGN-NLPFFYENFAGGYGS 660
Db 593 IPGSDNKYYKLSADVQGFYPLDRHRWVVSASAGYAN-GFGNKRLPFFQTYTAGGIGS 651
QY 661 VRGYDQSSLSGPRSQAYLTARRGQOT---TLGEVVGGNALATFGSELILPLPFKGDWT-DQ 716
Db 652 LRGFAYGSIGPNA---IVAEGHNGTENFKISSDVIGGNATITASAEILVPPFVSDKSONT 708
QY 717 VRPVIFIEGGGVFTT-GMDKQTDLTQFDKDPQATAEQNAKAANRPLLTDQKQLRYSGAV 775
Db 709 VRTSLFVDAASVWMTKWSKDGLESKVLKD-----LPDYGKSSRIRASTGV 755
QY 776 GATWYTPIGPLISISYAKPLNKKONDQTDVQFQIGSVF 813
Db 756 GFQWQSPIGPLVFSYAKPIKKYENDVDVEQFQSIGGSF 793

RESULT 4
D151_HAEIN STANDARD; PRT; 797 AA.
AC P46024;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-Ag) (Outer membrane protein D15).
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OC NCBI_TaxID=727;
RP [1]
RN SEQUENCE FROM N.A.
RC STRAIN-SEROTYPE B:
RX MEDLINE=95255676; PubMed=7737523;
RA Flack F.S., Loosmore S., Chong P., Thomas W.R.;
RT The sequencing of the 80-kDa D15 protective surface antigen of
RL Haemophilus influenzae."
RN Gene 156:97-99(1995).
RP [2]
RN SEQUENCE FROM N.A.
RC STRAIN-MINNA, AND EAGAN;
RX MEDLINE=97427952; PubMed=92841140;
RA Loosmore S.M., Yang Y.P., Coleman D.C., Shortreed J.M., England D.M.,
RA Klein M.H.;
RT "Outer membrane protein D15 is conserved among Haemophilus influenzae
RT species and may represent a universal protective antigen against
RT invasive disease."
RL Infect. Immun. 65:3701-3707(1997).
CC -|- SUBCELLULAR LOCATION: Outer membrane.
CC -|- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U13961; AAA85645.1; -
DR EMBL; U60832; AAB61974.1; -
DR EMBL; U60833; AAB61976.1; -
DR InterPro; IPR000184; Bac_surAg_D15.
DR Pfam; PF01103; Bac_surfaceAg; 1.
KW Antigen; Outer membrane; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 797 PROTECTIVE SURFACE ANTIGEN D15.
SQ SEQUENCE 797 AA; 87675 MW; 2F93DE538696AF1B CRC64;
```

```
Query Match 25.0% Score 1052.5 DB 1; Length 797;
Best Local Similarity 31.6% Pred No. 1.9e-54;
Matches 260; Conservative 160; Mismatches 333; Indels 69; Gaps 20;

QY 23 STHAAADFMANDITITGLQRVTIESLSQSVLFFRLGVVSENOLADGVKALYATGNFSVD 82
Db 14 TTTVFAAPFAKDIRVGVGDLEQQIRASLPVRAGQRTVDVANIVRSLSFVSGREDDV 73
QY 83 QVYHQEGRI-IYQTERPLIAEINFEGRNLLPKBGLQBLKAGLAVGQPLKQATVQMIE 141
Db 74 KA-HOEGDVLVSVVAKSIISDVKIKGNSVITPEALKQNLKDANGFKVGDVIREKLNFE 132
QY 142 TELTNOVISOCYVTEITVTKMLDGNRKVLDMTFAEGKPARVVDINTIGNQHFSDADLI 201
Db 133 KSVKEHYASGVRYNATPEVIVNTLPNNRAELIIQINEDDKRAKLASLFFKGNESVSSSTLQ 192
QY 202 DVLAIKONKINPLSKADRYTOEKLVTSLNLRARYLNAGFVFRFEIKDAKLININEDKNRIF 261
Db 193 EQMELQPDSSWWKL-WGNKFEQAQPEKDLQSIIRDYLLNNGYAKAQITKTQVLDNEKTKVN 251
QY 262 VELSLHEGEQYRFGQTOFLGNLTYTQAELEALLK-FRAEEGFSQAMLEQTTNNISTKFGD 320
Db 252 VTIDVNEGLQTDLSRSARIIGNLGMSAELEPLLSALHLNDTFRSDIADVENAIKALGE 311
QY 321 DGYVYAIQIRPVTRINDESRVVDVEYIDPVHPVYVVRINFTGNEKFTQDEVLRREMROLEG 380
Db 312 RGYGSATVNSVPDFDDANKTLAITLVVDAGRRLLTVRQLRPEGNIVSADSLRQEMRQEG 371
QY 381 ALASNQIKLSRARLMRTGTFKKHVTVDTR--PVPNSPDQVDVNVFVEQPSGSGSTIAAGY 438
Db 372 TWYNSQLVELGKIRLDRGTGFFE--TVENRIDPTINGSDENVDDVVYKVKERNTGSI 429
QY 439 SQSGGVTFQFDVSONNPMGTGKHVNASFSSRSETREVYSLGTMNPFYTVNGVQSLSGYR 498
Db 430 GTESGISIQASVQDNFLGTGAASVITAGTGYTSVNLGTYEFTYTKDGV--SLGNVF 487
QY 499 KTRYDNK--NISNYLDSYGGSLSYGYPIDENQIRISFGL-----NADNTKLHGGRFMGI 550
Db 488 FENYDNKSDTSSNYKRTTYGNSVNTLGFVPVNNNSYVGLGHTYKNSFALEYNRNLYI 547
QY 551 SNVKQLMADGGKIQVDNNGIPDFKHDDYTYNAILGNVSSLDPRVFPFQGMHSDV--LT 608
Db 548 QSMK-----FKNGIKTNDDEFD-----FCWNYNSLNRGYFPTKGVKASLGGRTV 592
QY 609 VGFEGDKTHOKVVYQGYNIYRPF-----IKKSVLRGVAKLGYGN-NLPFFYENFAGGYGS 660
Db 593 IPGSDNKYYKLSADVQGFYPLDRDLHWVVSASAGYAN-GFGNKRLPFFQTYTAGGIGS 651
QY 661 VRGYDQSSLSGPRSQAYLTARRGQOTTLG-----EVVGGNALATFGSELILPLPFKGDW 713
Db 652 LRGFAYGSIGPNA---IYAEYNGSGTGTPEKTISSDVIGGNATATASAEILVPTPFVSDK 708
QY 714 I-DOVRPVLIFIEGGGVFTT-GMDKQTDLTQFDKDPQATAEQNAKAANRPLLTDQKQLRY 771
Db 709 SQNTVTSIFVDAASVWMTKWSKDGLESVDLK-----RLPDYGKSSRIRA 755
QY 772 SAGVGTATWTPIGPLISISYAKPLNKKONDQTDVQFQIGSVF 813
Db 756 STGVGFQWQSPIGPLVFSYAKPIKKYENDVDVEQFQSIGGSF 797

RESULT 5
Y237_BUCAI STANDARD; PRT; 617 AA.
AC P57331;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein BU237 precursor.
GN BU237.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum symbiotic bacterium).
```



[illegible]

```

RESULT 6
YTFM_HAEIN          STANDARD;      PRT;    578 AA.
ID YTFM_HAEIN
P44038;
DC 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DR 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein HI0698 precursor.
DN HI0698.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OX Haemophilus
NCBI_TaxId=727;
[ ]
RN SEQUENCE FROM N.A.
RP STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Karlavage A.R., Sulton G.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton C., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Spriggs D.T., Saudke D.M., Brandon R.C.,
RA Fine L.D., Frichman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
NT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
[2]
RN IDENTIFICATION BY MASS SPECTROMETRY.
RX MEDLINE=20137488; PubMed=10675023;
RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,
RA Gray C., Fountoulakis M.;
RT "Two-dimensional map of the proteome of Haemophilus influenzae.";
RL Electrophoresis 21:411-429(2000).
CC -!- SIMILARITY: STRONG, TO E-COLI YTFM.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
-----
CC
CC EMBL; U32752; AAC22357.1; -.
DR TIGR; HI0698; -.
KW Signal; Complete proteome.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 578 PROTEIN HI0698.
SQ SEQUENCE 578 AA; 65812 MW; 34F9AC189C505876 CRC64;
Query Match 6.1%; Score 256.5; DB 1; Length 578;
Best Local Similarity 18.5%; Pred.No 6.5e-08;
Matches 132; Conservative 94; Mismatches 219; Indels 267; Gaps 25;
QY 152 GYYNTEITVKQTMDGNRVKLDMTFAEGCKPARV--VDINIIGNQHFSADLDIVLAIKDN 209
| | | : : : : : | : : | : | : | : | : | : | : |
DB 76 GYESSVRFERKQGRKDLLIAHVTGPTEKTAGTDVQIEG-----AAQDE 123
QY 210 KINPLSK-----ADRYTQEKLVTSLNLRAKYLNAGVFRETKAKLINEDKNRIF 261
| | | | | : : : : : | : | : | : | : | : | : | : | : |
DB 124 NFNALRNLPKGVLVBHQHYDDYDKTAISRLA---LNRGYFDGNFKISRLEISPETHQAW 180
QY 262 VEISLHEGEQYRFQGTQFLGNLTVTOAE----LEALLKFKAEGFSOAMLEQTNNIST 316
: : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 181 WRMLFDSGVRYHY-----GNITFSHSQIRDDYLNNLTINKSGDPY-----LMNWISD 227
QY 317 KFGD--DGYTYAQTRPVTRINDSRSTVDVEYYIDPVHPVTVRRINTGFNKETOEVLRRE 374
| : : : : : : : : : | : | : | : | : | : | : | : | : |

```



QY 676 YLTARRGOQTTLGVEVVGNGNALTATGSELILPLPFKGDWIDQVRPVIFIEGGQVFTTGM 735  
 Db 488 YAN-----GDLKASKLIT--GSLEYQYNVTKW-----WGAVFVDSGEA 525  
 QY 736 KOTIDLTOKFOATAEQNAKAAANRPLLTQDKQLRYSGAGVATWTPICPLSLISYAKPLN 795  
 Db 526 VSDIRSDPKT-----GTVGVWRWESPVPKIDFAVPA 560  
 QY 796 KQNDQDTDTVOFQIG 810  
 Db 561 DKDE--HGLQFYIG 572  
 RESULT 8  
 CG51\_HUMAN  
 ID CG51\_HUMAN STANDARD; PRT; 469 AA.  
 AC Q9Y512; OSNW85; O9UC0M9;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DI 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Protein CGI-51.  
 GN CGI51.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20057165; PubMed=10591208;  
 RA Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,  
 RA Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.,  
 RA Baguley C., Bailey J., Barlow K., Bates K.N., Beasley O., Bird C.P.,  
 RA Blakey S., Bridgeman A.M., Buck D., Burgess J., Burrill W.D.,  
 RA Burton J., Carder C., Carter N.P., Chen Y., Clark G., Clegg S.M.,  
 RA Cobley V., Cole C.G., Collier R.E., Connor R.E., Conroy D., Corby N.,  
 RA Coville G.J., Cox A.V., Davies J., Dawson E., Dhami P.D., Dockree C.,  
 RA Dodsworth S.J., Durbin R.M., Ellington A., Evans K.L., Fey J.M.,  
 RA Fleming K., French L., Garner A.A., Gilbert J.G.R., Goward M.E.,  
 RA Graham D., Griffiths M.N., Hall C., Hall R., Hall-Tamlyn G.,  
 RA Heathcote R.W., Ho S., Holmes S., Hunt S.E., Jones M.C., Kershaw J.,  
 RA Kimberley A., King A., Laird G.K., Langford C.F., Leversha M.A.,  
 RA Lloyd C., Lloyd D.M., Martyn I.D., Mashreghi-Mohammadi M.,  
 RA Matthews L., Mccann O.T., Mcclay J., McLaren S., McMurray A.,  
 RA Milne S.A., Mortimore B.J., Odeh C.N., Pavitt R., Pearce A.V.,  
 RA Pearson D., Phillimore B.J., Phillips S.H., Plumb R.W., Ramsay H.,  
 RA Ramsey Y., Rogers L., Ross M.T., Scott C.E., Sehra H.K., Skuce C.D.,  
 RA Smalley S., Smith M.L., Soderlund C., Spragon L., Steward C.A.,  
 RA Sulston J.E., Swann R.M., Vaudin M., Wall M., Wallis J.M.,  
 RA Whiteley M.N., Willey D., Williams L., Williams S., Williamson H.,  
 RA Wilmer T.E., Wilming L., Wright C.L., Hubbard T., Bentley D.R.,  
 RA Beck S., Rogers J., Shimizu N., Minoshima S., Kawasaki K., Sasaki T.,  
 RA Asakawa S., Kudoh J., Shintani A., Shibuya K., Yoshizaki Y., Aoki N.,  
 RA Mitsuyama S., Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S.,  
 RA Do A., Do T., Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S.,  
 RA Lai H., Lao H.I., Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E.,  
 RA Nguyen T., Pan H., Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S.,  
 RA Sloan D., Song L., Wang Q., Wang Y., Wang Z., White J., Willingham D.,  
 RA Wu H., Yao Z., Zhan M., Zhang G., Chissoe S., Murray J., Miller N.,  
 RA Minx P., Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H.,  
 RA Bourne S., Cordes M., Du Z., Fulton L., Goela D., Graves T.,  
 RA Hawkins J., Hinds K., Kemp K., Latreille P., Layman D., Ozersky P.,  
 RA Rohlfing T., Scheet P., Walker C., Wamsley A., Wohlmann P., Pepin K.,  
 RA Nelson J., Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R.,  
 RA Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,  
 RA Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,  
 RA Edlmann L., Kim U.J., Shizuya H., Simon M.I., Dumanaki J.P.,  
 RA Peyrard M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,  
 RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,  
 RA Khan A.S., Lane L., Tilahun Y., Wright H.;  
 RT "The DNA sequence of human chromosome 22.";  
 RL Nature 402:489-495(1999).  
 RP SEQUENCE FROM N.A.

RX MEDLINE=20272150; PubMed=10810093;  
 RT Lai C.-H., Chou C.-Y., Ch'ang L.-Y., Liu C.-S., Lin W.-C.;  
 RA "Identification of novel human genes evolutionarily conserved in  
 RT Caenorhabditis elegans by comparative proteomics.";  
 RL Genome Res. 10:703-713(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RA Isoqai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,  
 RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,  
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,  
 RA Ninomiya K., Iwavanagi T.;  
 RT "NEDO human cDNA sequencing project.";  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE UPF0140 (CGI-51) FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; AL035398; CAB51401.1; -;  
 DR EMBL; AK001087; AAD91498.1; -;  
 DR EMBL; AF151809; BAA34046.1; -;  
 FT CONFLICT 110 110 D -> G (IN REF. 3).  
 FT CONFLICT 345 345 I -> V (IN REF. 3).  
 FT CONFLICT 368 372 WAGGL -> LGRRW (IN REF. 2).  
 SQ SEQUENCE 469 AA; 51976 MW; 4F686FB7ACB08DEF CRC64;  
 Query Match 4.7%; Score 197.5; DB 1; Length 469;  
 Best Local Similarity 22.0%; Pred. No. 0.00014;  
 Matches 114; Conservative 82; Mismatches 212; Indels 111; Gaps 22;  
 QY 334 INDESRVDVE---YYIDPVHPVYRRINFTGNFKTQDEVLRRMQL---EGALASNQ 386  
 Db 22 LGEEAEFEVEPEAKOEILENKDVVQVHVHFDGLGRTKDDIIICEIGDFKAKNLIEVVR 81  
 QY 387 KTQLSRARLMRTGFEKHTV---DTRVPVNS-PDQVDVNFVVE--OPSGSSTTAAGYSOS 441  
 Db 82 KSHAREKLLRIGIFQVDVLDTCQGDALPGLDVTFTVTLRLTSGSYNTWVGNNEG 141  
 QY 442 GGVTFQFQVSNQNFMTGTHKVNASFSTRETVYSLGMTNP-----YFTVNGVSQSLS 494  
 Db 142 SMV---LGLKLPNLLGRAEKVTFQFSYGTSETSYGLSFEKPRPGNFRNFSVN--LYKVT 196  
 QY 495 GYV---RKTXYDNKNISNYVLDVSGGSLSYGYPIDENQRIISFGLNADNTKLHGFRWG 549  
 Db 197 GQFPSSSLRET-----DRGMAEYSFPIWKTS-----HTVKWEG 230  
 QY 550 ISNVKQLMADGGKIQVDNNGIPDFKHDTYTYNAILGNWYSSLDPRVFTPQGMSSHVDLT 609  
 Db 231 VVRELGLSLRTASFVAKESGHSLSKSLSHAMVI-----DSRNSSTLPRGALLKYNQEL 285  
 QY 610 -GF--GDKTHQKVVOGNIYRPFIKKSVLRGAKYGNLNP-----FYENFYAGGYG 659  
 Db 286 AGYTGQDVSFIEDFELQNLKLFDSVFS--ASFWGMGLVPITGDKPSSSTADRFYLGPT 343  
 QY 660 SVRGVQDSSIGPRSQAYLTARRQQQTTLGVEVVGNGNALTATGSELILPLPEK---GDWIDQ 716  
 Db 344 SIRGFSMHSIGQSE-----GDYLGGEAYWAGGLHLYTLPFRPGQGGFGE 390  
 QY 717 VRPVTFIEGGQVFDATGMDKQTDITLQFKDQATQNAKAAANRPLLTQDKQLRYSGAGV 776  
 Db 391 FTTFELNAGNL-----CNLNYGGPKAHIRKLAEC-----IRWSYGAG 429  
 QY 777 ATW-YTPIGPLSIYAKPLNKKNDQ--TDTVQFQIGSVF 813

430 IVLRGLNRIARELNYCVPMGVGTGDRICDGVQFGAIRF 458

Db

SEQUENCE 463 AA; 51427 MW; 6AABDDADCC5F2C66 CRC64;

Query Match 4.1%; Score 171; DB 1; Length 463;  
Best Local Similarity 23.4%; Pred. No. 0.005;  
Matches 98; Conservative 69; Mismatches 172; Indels 80; Gaps 22;

QY 336 DESRTVDVEYIDPVHVVYRRINFTGFKTQDEVLRLRMEQLEGAL-ASN-QKIQL--- 390  
Db 7 DGGASKDSKYDLSKI-SARVDRVNVSGLLRTHNDYV---MRAADGLFKASNFQDLMEAM 62  
QY 391 -SPEARLMRTGFFPKHVTVD---TRPVNPDQVDVNVFVVEQPSGSSITAAAGYSOGGVTF 446  
Db 63 STKSYLHEIGIFKDSVHIDVSRGADASPQGYEVTFKGNEMSRMGSGAGTEIGQNEG-SL 121  
QY 447 QFDVSONNPMGTGKHNASFSRSETR-EVYSLGTMNPYFTV---NGVSQSLSGYVYRKT 501  
Db 122 RTELTIPNILGRGENISLGSYSSTRANDLQLKFWKPFHFFKVENRPMESFS-IFRQT- 179  
QY 502 YDNKNISNVYLSYSGSLSYGPIDENQRIISFGLNAD-----NTKLHGGRFMGINSVKQL- 556  
Db 180 -DRFDISSF-----QTTNIGYLVDFSAHTMVGVDVSLKILQSKLN---FLAIYFNFLT 229  
QY 557 -----MADGGKIQVDNNGIPDKHY-----TTYNAILGWNYSYSLDRPVFPTQG 600  
Db 230 HSLQYENAIRDVGLL---NKSVPFAIRDHCGPKLASLLRYSV---YDRDGNVFFPRG 282  
QY 601 M-SHSVDLTVGF-GDKTHQKVYQGNYYRPFKKSVLRGAKLGY-----GNLPPFYEY 652  
Db 283 IYLSKVNEYCGLGNNVAYTSSTAHELNVPLFAGLVAQFCARVGVKVKETKNTTQLPISSL 342  
QY 653 FYAGGYSVGRGVDQSSGLGRSQAVLTARRGQTTLGEVGVGNALATSGSELILPLPKG 711  
Db 343 FYCGGFLIRLKFKGAGPVVES-----TPIGAQSFWCTGAHLNAPLPFAG 388

RESULT 10  
HLVA\_PROMI  
ID HLVA\_PROMI STANDARD; PRT; 1577 AA.  
AC P16466;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 01-NOV-1990 (Rel. 16, Last annotation update)  
DE Hemolysin precursor.  
GN HPMA.  
OS Proteus mirabilis.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OX Proteus.  
OC NCBI\_TaxID=584;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 30-43.  
RC STRAIN-ISOLATE 477-12;  
RX MEDLINE=90170827; PubMed=2407716;  
RA Uphoff T.S., Welch R.A.;  
RT "Nucleotide sequencing of the Proteus mirabilis calcium-independent hemolysin genes (hpmA and hpmB) reveals sequence similarity with the Serratia marcescens hemolysin genes (shlA and shlB).";  
RL J. Bacteriol. 172:1206-1216(1990).  
CC -!- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY DEFINED.  
CC -!- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. HPMA REQUIRES HPMB FUNCTION.  
CC -!- SUBCELLULAR LOCATION: Outer membrane.  
CC -!- MISCELLANEOUS: THE CONSERVED AMPHIPATHIC DOMAINS IN SHLA AND HPMA MAY BE RESPONSIBLE FOR PORE FORMATION.  
CC -!- SIMILARITY: TO S. MARCESCENS HEMOLYSIN (SHLA).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

SEQUENCE FROM N.A.  
STRAIN=BERKELEY;  
MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F., George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D., Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G., Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D., Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S., Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P., Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W., Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C., Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M., Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Svirskas R., Tector C., Turner K., Venter E., Wang A.H., Wang X., Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
CC -!- SIMILARITY: BELONGS TO THE UPF0140 (CG7-51) FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AE003812; AAF58175.1; -  
DR FlyBase; FBgn0033989; CG7639.  
CC Hypothetical protein.

3.7%; Score 154.5; DB 1; Length 1256;

Best Local Similarity 19.0%; Pred. No. 0.19;  
Matches 199; Conservative 125; Mismatches 319; Indels 405; Gaps 55;

QY	9	POVSAMTAVVMVMSHTHAQAADFMANDITITGQRVTIESQSQSVLPRFGQVVSENL---	66
Db	25	FGAASVLLGVSLVLGAGAQV---VKADETVAS-SEPTIAS--SVAPASTEVAEEAEKTN 78	
QY	67	ADGVKALYANGNFSDVQVYHQEGRIIYQVT-ERPLTAEINFENRRLIPK---EGLOGEGL 122	
Db	79	AENTSAVATTS--TEVE---KAKAVLEQVTSPLAGL---GQKELAKEDATLAKAIE 130	
QY	123	NA--GLAVGOPL---KOATYOMITETL-----NOYISQVYNTETI----- 158	
Db	131	DAQTKLAAAKAILADSEATVEQVEAQVAAVKVANEALGNELQYTVDGLLTAALDTVAPD 190	
QY	159	TVKOTML--DGNRVKLDMT-----FAEGKPARVDVINIIGNOHFSDADLIDVLAIKDNK 210	
Db	191	TTASTLKGEGEGTLLDSTTTATPSSMAEPNGAAIAP-----HTLRTQDG- 234	
QY	211	INPLSKADRYTOEK--LVTSLENLRAK-YLNA---GFVRPEIKD-----AKLNINED 256	
Db	235	IKATSEPNWYTFESYDLYSNKNMASTYKGAEDAIIRYSLDNDSSTTAVLAEL--VSRT 293	
QY	257	KNRIFVISISHEGEQYRFGOTEL---GNLTYT-----QAELEALLKFAEGEFSQAM 306	
Db	294	TGDLVKYITPEGESYTFSHPTKVANNANSITVYDTSLASANTPCALKFSANDOVYSTI 353	
QY	307	L-----EOTTNNISTKFGD-----DGYYY----- 325	
Db	354	IVPAYQINTTRYTESGKVLATYGLQTIAGQVVPSPSVRFYTDYVATTTKRAVGQPYK 413	
QY	326	-----AQRPVTRINDESRTVDVEYIDVPHPVYVRR.INFTGNKFTQDEVLR 372	
Db	414	GTVVLACTVQKDVQVKVIREIVENDQAVLKFYILD---PTYKGEVDWRGT----- 461	
QY	373	REMROLEGALASNQIKLSRLARLMRTFFKFKHVPVTRPVPNSPDQVDVNFVVEEQPSGS 432	
Db	462	-----DTTGFIELLT-----TSPTTYKVGTYD----- 484	
QY	433	TIAGYSQSQGVTFQFDV-----SQNNFMGTGKHVNAPSRSTR-----EYV- 475	
Db	485	-----YNINSKITAPETIDPTKNVMVFKESEQNEQSKYRVIAQWSGDETTKGIYKGYI 539	
QY	476	-----SLGMNPFTVNGVSQSLSG-----YYRK 499	
Db	540	ATQVWTTKLTNEGWEDYSDDOAGIKFNKGFWPAGVQNTURNATPATAVETVYIKES 599	
QY	500	TKY-----DNKNISNYLDSYGGSLGYCYPIDENQRISFGLNADNT----- 540	
Db	600	SKYGDVIVEYDYGKQIVNSVVDTPKPSALGTEYNTDVRRPASLVAADGTVYFKEYKS 659	
QY	541	---KLHGGRPMGISNVKQLMADGGKTQ---VDNNG-----IPDPKHDTYTTNAILGNWY 588	
Db	660	DSAKTTGTVVVAGTTTYKYVYKAGSVNVNFVDINGKVIKAPVSDEK-----DAKPGYNY 713	
QY	589	SSLDRPVFPTOGMSHSDVLTVGFQDKTHQKVYVVOGNIYRPFKIKKSVLRG---YAKLGYCN 645	
Db	714	DT-----DL-----DOKLASTIFEGKEY-----KLVPAGDYDVGKVGKN 748	
QY	646	NLPFYENFYAGG-----YGSVRG-----YDOS-----SL 669	
Db	749	NLIEVCNNTAKGIDPTTGKIEAGVNKEVTVYVRAVTGVSVVVYKDEGNVIKDPTDYSD 808	
QY	670	GPRSQAALT-----ARRGOITL-----GEWCGNALATFGESELLILPLPFK 710	
Db	809	APVGDAYTTDDKKPNEIITKDGSRVYLVPSKTDGEENGKVIETGTTVTVYVQV----- 862	
QY	711	GDWIDQVRPVIFTEGGQV---EDTTGMQRQITDLQFKDPPQATBQNAKAANRPLLTQDK 767	
Db	863	ANWIPPEPNVPETDRPKVYPPEPDEPD--EPIDPT-----TPGTNGVEVPNIP----- 908	
QY	768	QLRYSAGVGATWYTPIGPLSISYAKPLN 795	

Db 909 ---YVPG-----YTPVDPKDNTPLKPID 928

```

RESULT 12
F3P_LACLC
ID   P3F_LACLC          STANDARD;      PRT;   1902 AA.
AC   P15292;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   PIII-type proteinase precursor (EC 3.4.21.96) (Lactocepin) (Cell wall-
DE associated serine proteinase).
GN   prnp.
OS   Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OG   Plasmid.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Lactococcus
OX   NCBI_TaxID=1359;
[1]
RN   SEQUENCE FROM N.A., AND SEQUENCE OF 188-197.
STRAIN=SK11;
MEDLINE=89340435; PubMed=2760036;
RA   Vos P., Simons G., Siezen R.J., de Vos W.M.;
RT   "Primary structure and organization of the gene for a procaryotic,
RT cell envelope-located serine proteinase.";
RL   J. Biol. Chem. 264:13579-13585(1989).
CC   -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC GROWTH OF THE BACTERIA ON MILK.
CC   -!- CATALYTIC ACTIVITY: Endopeptidase activity with very broad
CC specificity, although some subsite preference have been noted,
CC e.g. large hydrophobic residues in the p1 and p4 positions, and
CC pro in the p2 position. Best known for its action on caseins,
CC although it has been shown to hydrolyse hemoglobin and oxidized
CC insulin B-chain.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR   EMBL; J04962; AAA03533.1; ALT_SEQ.
DR   PIR; A32634; A32634.
DR   HSSP; P00782; ZSBT.
DR   MEROPS; S08.019; -.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR003137; PA.
DR   InterPro; IPR000209; Peptidase_S8.
DR   Pfam; PF00746; Gram_pos_anchor; 1.
DR   Pfam; PF02225; PA; 1.
DR   Pfam; PF00082; Peptidase_S8; 3.
DR   PRINTS; PR00723; SUBTILISIN.
DR   PROSITE; PS00136; SUBTILASE_ASP; 1.
DR   PROSITE; PS00137; SUBTILASE_HIS; 1.
DR   PROSITE; PS00138; SUBTILASE_SER; 1.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; 1.
KW   Hydrolase; Serine protease; Cell wall; Zymogen; Signal; Plasmid;
transmembrane.
FT   SIGNAL             1
FT   PROPPEP           34
FT   CHAIN              188
FT   DOMAIN             188
FT   TRANSMEM           1877
FT   DOMAIN             1896
FT   ACT_SITE           217
FT   ACT_SITE           281
FT   ACT_SITE           620
FT   DOMAIN            1867
FT   SIGNAL             33
FT   PROPPEP           34
FT   CHAIN              188
FT   DOMAIN             188
FT   TRANSMEM           1877
FT   DOMAIN             1896
FT   ACT_SITE           217
FT   ACT_SITE           281
FT   ACT_SITE           620
FT   DOMAIN            1867
FT   SIGNAL             33
FT   PROPPEP           34
FT   CHAIN              188
FT   DOMAIN             188
FT   TRANSMEM           1877
FT   DOMAIN             1896
FT   ACT_SITE           217
FT   ACT_SITE           281
FT   ACT_SITE           620
FT   DOMAIN            1867
FT   SIGNAL             33
FT   PROPPEP           34
FT   CHAIN              188
FT   DOMAIN             188
FT   TRANSMEM           1877
FT   DOMAIN             1896
FT   ACT_SITE           217
FT   ACT_SITE           281
FT   ACT_SITE           620
FT   DOMAIN            1867
FT   SIGNAL             33
FT   PROPPEP           34
FT   CHAIN              188
FT   DOMAIN             188
FT   TRANSMEM           1877
FT   DOMAIN             1896
FT   ACT_SITE           217
FT   ACT_SITE           281
FT   ACT_SITE           620
FT   DOMAIN            1867
FT   SIGNAL             33
FT   PROPPEP           34
FT   CHAIN              188
FT   DOMAIN             188
FT   TRANSMEM           1877
FT   DOMAIN             1896
FT   ACT_SITE           217
FT   ACT_SITE           281
FT   ACT_SITE           620
FT   DOMAIN            1867
FT   SIGNAL             33
FT   PROPPEP           34
FT   CHAIN              188
FT   DOMAIN             188
FT   TRANSMEM           1877
FT   DOMAIN             1896
FT   ACT_SITE           217
FT   ACT_SITE           281
FT   ACT_SITE           620
FT   DOMAIN            1867
FT   SIGNAL             33
FT   PROPPEP           34
FT   CHAIN              188
FT   DOMAIN             188
FT   TRANSMEM           1877
FT   DOMAIN             1896
FT   ACT_SITE           217
FT   ACT_SITE           281
FT   ACT_SITE           620
FT   DOMAIN            1867
FT   SIGNAL             33
FT   PROPPEP           34
FT   CHAIN              188
FT   DOMAIN             188
FT   TRANSMEM           1877
FT   DOMAIN             1896
FT   ACT_SITE           217
FT   ACT_SITE           281
FT   ACT_SITE           620
FT   DOMAIN            1867
FT   SIGNAL             33
FT   PROPPEP           34
FT   CHAIN              188
FT   DOMAIN             188
FT   TRANSMEM           1877
FT   DOMAIN             1896
FT   ACT_SITE           217
FT   ACT_SITE           281
FT   ACT_SITE           620
FT   DOMAIN            1867
FT   SIGNAL             33
FT   PROPPEP           34
FT   CHAIN              188
FT   DOMAIN             188
FT   TRANSMEM           1877
FT   DOMAIN             1896
FT   ACT_SITE           217
FT   ACT_SITE           281
FT   ACT_SITE           620
FT   DOMAIN            1867
FT   SIGNAL             33
FT   PROPPEP           34
FT   CHAIN              188
FT   DOMAIN             188
FT   TRANSMEM           1877
FT   DOMAIN             1896
FT   ACT_SITE           217
FT   ACT_SITE           281
FT   ACT_SITE           620
FT   DOMAIN            1867
FT   SIGNAL             33
FT   PROPPEP           34
FT   CHAIN              188
FT   DOMAIN             188
FT   TRANSMEM           1877
FT   DOMAIN             1896
FT   ACT_SITE           217
FT   ACT_SITE           281
FT   ACT_SITE           620
FT   DOMAIN            1867
FT   SIGNAL             33
FT   PROPPEP           34
FT   CHAIN              188
FT   DOMAIN             188
FT   TRANSMEM           1877
FT   DOMAIN             1896
FT   ACT_SITE           217
FT   ACT_SITE           281
FT   ACT_SITE           620
FT   DOMAIN            1867
FT   SIGNAL             33
FT   PROPPEP           34
FT   CHAIN              188
FT   DOMAIN             188
FT   TRANSMEM           1877
FT   DOMAIN             1896
FT   ACT_SITE           217
FT   ACT_SITE           281
FT   ACT_SITE           620
FT   DOMAIN            1867
FT   SIGNAL             33
FT   PROPPEP           34
FT   CHAIN              188
FT   DOMAIN             188
FT   TRANSMEM           1877
FT   DOMAIN             1896
FT   ACT_SITE           217
FT   ACT_SITE           281
FT   ACT_SITE           620
FT   DOMAIN            1867
FT   SIGNAL             33
FT   PROPPEP           34
FT   CHAIN              188
FT   DOMAIN             188
FT   TRANSMEM           1877
FT   DOMAIN             1896
FT   ACT_SITE           217
FT   ACT_SITE           281
FT   ACT_SITE           620
FT   DOMAIN            1867
FT   SIGNAL             33
FT   PROPPEP           34
FT   CHAIN              188
FT   DOMAIN             188
FT   TRANSMEM           1877
FT   DOMAIN             1896
FT   ACT_SITE           217
FT   ACT_SITE           281
FT   ACT_SITE           620
FT   DOMAIN            1867
FT   SIGNAL             33
FT   PROPPEP           34
FT   CHAIN              188
FT   DOMAIN             188
FT   TRANSMEM           1877
FT   DOMAIN             1896
FT   ACT_SITE           217
FT   ACT_SITE           281
FT   ACT_SITE           620
FT   DOMAIN            1867
FT   SIGNAL             33
FT   PROPPEP           34
FT   CHAIN              188
FT   DOMAIN             188
FT   TRANSMEM           1877
FT   DOMAIN             1896
FT   ACT_SITE           217
FT   ACT_SITE           281
FT   ACT_SITE           620
FT   DOMAIN            1867
FT   SIGNAL             33
FT   PROPPEP           34
FT   CHAIN              188
FT   DOMAIN             188
FT   TRANSMEM           1877
FT   DOMAIN             1896
FT   ACT_SITE           217
FT   ACT_SITE           281
FT   ACT_SITE           620
FT   DOMAIN            1867
FT   SIGNAL             33
FT   PROPPEP           34
FT   CHAIN              188
FT   DOMAIN             188
FT   TRANSMEM           1877
FT   DOMAIN             1896
FT   ACT_SITE           217
FT   ACT_SITE           281
FT   ACT_SITE           620
FT   DOMAIN            1867
FT   SIGNAL             33
FT   PROPPEP           34
FT   CHAIN              188
FT   DOMAIN             188
FT   TRANSMEM           1877
FT   DOMAIN             1896
FT   ACT_SITE           217
FT   ACT_SITE           281
FT   ACT_SITE           620
FT   DOMAIN            1867
FT   SIGNAL             33
FT   PROPPEP           34
FT   CHAIN              188
FT   DOMAIN             188
FT   TRANSMEM           1877
FT   DOMAIN             1896
FT   ACT_SITE           217
FT   ACT_SITE           281
FT   ACT_SITE           620
FT   DOMAIN            1867
FT   SIGNAL             33
FT   PROPPEP           34
FT   CHAIN              188
FT   DOMAIN             188
FT   TRANSMEM           1877
FT   DOMAIN             1896
FT   ACT_SITE           217
FT   ACT_SITE           281
FT   ACT_SITE           620
FT   DOMAIN            1867
FT   SIGNAL             33
FT   PROPPEP           34
FT   CHAIN              188
FT   DOMAIN             188
FT   TRANSMEM           1877
FT   DOMAIN             1896
FT   ACT_SITE           217
FT   ACT_SITE           281
FT   ACT_SITE           620
FT   DOMAIN            1867
FT   SIGNAL             33
FT   PROPPEP           34
FT   CHAIN              188
FT   DOMAIN             188
FT   TRANSMEM           1877
FT   DOMAIN             1896
FT   ACT_SITE           217
FT   ACT_SITE           281
FT   ACT_SITE           620
FT   DOMAIN            1867
FT   SIGNAL             33
FT   PROPPEP           34
FT   CHAIN              188
FT   DOMAIN             188
FT   TRANSMEM           1877
FT   DOMAIN             1896
FT   ACT_SITE           217
FT   ACT_SITE           281
FT   ACT_SITE           620
FT   DOMAIN            1867
FT   SIGNAL             33
FT   PROPPEP           34
FT   CHAIN              188
FT   DOMAIN             188
FT   TRANSMEM           1877
FT   DOMAIN             1896
FT   ACT_SITE           217
FT   ACT_SITE           281
FT   ACT_SITE           620
FT   DOMAIN            1867
FT   SIGNAL             33
FT   PROPPEP           34
FT   CHAIN              188
FT   DOMAIN             188
FT   TRANSMEM           1877
FT   DOMAIN             1896
FT   ACT_SITE           217
FT   ACT_SITE           281
FT   ACT_SITE           620
FT   DOMAIN            1867
FT   SIGNAL             33
FT   PROPPEP           34
FT   CHAIN              188
FT   DOMAIN             188
FT   TRANSMEM           1877
FT   DOMAIN             1896
FT   ACT_SITE           217
FT   ACT_SITE           281
FT   ACT_SITE           620
FT   DOMAIN            1867
FT   SIGNAL            
```

FT SEQUENCE 1902 AA; 200550 MW; 87CECBAA9345F9D3 CRC64; PROTEINS.  
SQ  
Query Match 3.5%; Score 149; DB 1; Length 1902;  
Best Local Similarity 19.6%; Pred. No. 0.74;  
Matches 175; Conservative 116; Mismatches 356; Indels 246; Gaps 44;  
QY 4 SYFKGQVQSMAYMVMV-MSTHAQAADPMANDIIT-----GLQRTIESLQSVLPFRUG 58  
DB 4 YKQKLTALDELKVTENMTAQPINLNNVIVPRQAGLVQVKAADALEKNPS 705  
QY 59 QVVSNQIADGKALYATGNFSDVYHQ--EGRIYQVTERPLIAEINFEGRNLI-----112  
DB 706 TVVAEN---GYPAVELKDTSTDKLTFTNRTHETLYQ---MDSNTDNAVYTSAT 758  
QY 113 -PKEGL--QEGKKNAGLAVGQPLKQATQMIEETL-----NOYISQGYNTEITVK 161  
DB 759 DPNQSVGLVDKKIDGAIRAGSNITVPAGKTAQIEFTLSLPKSFDDQOQFVEGFLNFKGS--816  
QY 162 QTMLDGNRVKLDMT--FAEGKPARVVDINIIGNHFSDADLIDVLAIKDNKINPLSKADR 219  
DB 817 ---DGRNLNLYMGFGDNDGKIVD-SLNGITYSPAGNFGTVPLLKNK-----NTGTQ 867  
QY 220 YTOEKLVTSLNLRKAYLNAQVREIKDAKUNEDKNRIEVSILH-----267  
DB 868 Y-YGMVTDADGNKT-----VDDQAIASFSSDKNALYNDISMKYILLRNISNVQV 915  
QY 268 -----EGEQYRFGQTFGLNLTYYTOAELEALLKFAEGFSQAMLEFOTNNISTKFGDDG 322  
DB 916 DILDQGNKVTTLSSSTNRKKTYYNAHQOYIYYNA-PAWDGTYDQDRDGNIKT--ADDG 972  
QY 323 YYYAQRVPTVINDESRTVDVEYYIDPVHPVYVRINF---TGNEFKTQDEVLRRERMOLE 379  
DB 973 SYTYRISGVPEGGRQVDFPFKLDKAPT-VRHVALSAKTENGKTQYLLTAEKADDL 1031  
QY 380 GALASNQIKLSRRLMRTGFFKHVTVTRPVNPSDQVDVNFVVEQPSGSTIAAGYS 439  
DB 1032 GLDPT-----KSVKTEINEVTN-----LDATF-----TDAGTTADGVT 1064  
QY 440 QSGGVTFQDVSONFMGTGKHV-----NASFSRSETREYVSLGTMNPTVTVNG-----488  
DB 1065 K-----IETPLSDQAALGNDNSAELYTDNASNATDQASVKPGSTSFDLIVNGGGIP 1121  
QY 489 ---VSQSLSGYRKTK-----YDNKNISNYVLDVSGYSGLSYGY 524  
DB 1122 DKISSTTGYEANTOGGTYTFSGTYPAAVDGYTDAQCKKHDLTNTYDAATNSFTASMP 1181  
QY 525 I-DENQISFGLNADNTKLHGGRFMGISNVKOLMADGKIQVDNNGIPDFKH-----575  
DB 1182 VTNADYAAQVDLYAD--KAHTQLLKHFDTKVRMLMAPFTDLKFNNGSDQTSATIKVTGT 1239  
QY 576 ---DYTTNAILGWYSISLDRPVFTQGMH-SVDLTGVFGDK-----THQ 617  
DB 1240 VSADTKTVN--VGHVFAALD-----AOHHSVDVPVNYGNTIKVTATDKDGNNTTQ 1290  
QY 618 KVVYQGNIRP-FIKKSVLRGYAKLGYGNLPPFYENFYAGGYSVGRVDQSSLSGPRSOAY 676  
DB 1291 KTIITSS--YDPLMKSVT-----FDQGVKFGTNKF--NATSAKFYDPKT---GIAT 1335  
QY 677 LTARGQQTTEGVVGGNA-----LATFGELEILPLPF-----KGMW-----713  
DB 1336 ITGKVKHPTTTLQVDGKQIPIKDDLTFTSFTLDLGLGQK-----PFGVVGVDTTQNKTFQ 1390  
QY 714 -----IDQVRVIEGGQVDTTGMDKQITDLTQFKDPQ-----ATAEQNAK 756  
DB 1391 EALSFILOAVP-----TSLSDSSDAPVYTNPNQITGTATDQNAO 1432  
RESULT 13  
YDF6\_SCHPO  
ID YDF6\_SCHPO STANDARD; PRT; 475 AA.  
AC Q10478;

DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical 51.8 kDa protein Cl7C9.06 in chromosome I.  
GN SPAC17C9.06.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OX Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RA Murphy L., McDougall R., Jones L., Simpson I., McNeil A., Harris D.,  
RA Bartell B.G., Rajandream M.A., Walsh S.V.,  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: BELONGS TO THE UPF0140 (Cti-51) FAMILY.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; 273099; CAA97352.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 475 AA; 51762 MW; 8FEC5B5FF4B5DE15 CRC64;  
Query Match 3.5%; Score 148; DB 1; Length 475;  
Best Local Similarity 19.1%; Pred. No. 0.12;  
Matches 102; Conservative 96; Mismatches 209; Indels 126; Gaps 24;  
QY 315 STKFGDDGYYYAQIRPVTRINDESRTVDVEYY-----IDPVHPVYVRINFETGNFT 366  
DB 7 STSPSD-----IPAVNEESKLSAETFKSLSEILAENSTLPVGSIIRVTGAHT 57  
QY 367 QDEVLRREMR-----OLEGALASNQIKLSRRLMRTGFFKHVTVTRPVNS---414  
DB 58 RPSFIRKVLCTCLDTSKPAKRSLLLETNAIOETTNLMFNVYETANTKIDRASSVSG 117  
QY 415 PQQVDVNFVVEBP-----SGSSTIAAGYSQSGV--TFQDVSONFMGTGKHVNASFS 467  
DB 118 DDDLDTVIOVKEKPRLYVETGTV-----GNVEGNHANVLARNVFGAELLGNSV 169  
QY 468 -RSETREYVSLGTMNPDYFTVNGVVSQSLGYRRTKYDNKNISNYVLDVSGYSGLSYGY 526  
DB 170 YGTRNRSTMSVNFETPVNADPKTRLRFNGH--SNLRDNKSISSHDLTLTKGITLSL---222  
QY 527 ENQISFGLNADNTKLHGGRFMGISNVKOLMADGKIQVDNNGIPDFKHDTTYTNAILGW 586  
DB 223 QHODLWSGHHLSQNL---LWRQVTHLTAYASPSVRLAEGDSLKQSLSYTYTR-----272  
QY 587 NTSSSLDPRVFPFQGG--MHSVSLDT-VGF--GDKTHQKVYQGVNIYRPFKKSV-----634  
DB 273 --DTRDHLMIPTKGDVVRQTLAELAGFLPGDASFLKSEFWG-----QKVALNSSRS 323  
QY 635 --LRGYAKLGYGNL-----PFYENFYAGYGSVRYGQSSLSGPRSOAYLTARRGQO 687  
DB 324 VSLISARIGALHSLNKKQVSLCDREMLGGSTSLRGFSEDRIGPKD-----GRDS--373  
QY 688 GEVVGGNALATGSELIPLPFKGMWIDQVRPVIFTEGGQVDTTGMDKQITDLTQFKDP 747  
DB 374 ---LGGTAYMAFSMSLLFPLP-KVDASKPFRQLQLFANAGGLSNLTS-----PNP 418  
QY 748 QATAEQNAKANRPLLTQDKQLRYSAGVGATWYTPIGPLSISYAKPLNKKQND 800  
DB 419 CGTYK---SILSKPCI-----STGLGLVYATPAARFELNFTLPIATEKD 460  
RESULT 14



```

YEHF_ECOLI
ID YEHF_ECOLI STANDARD; PRT; 826 AA.
AC P33341;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical outer membrane usher protein yehB precursor.
GN YEHF OR B2109.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / BHB2600;
RA Richterich P., Lakey N., Gryan G., Jaehn L., Mintz L., Robison K.,
RL Church G.M.;
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12."
RT Science 277:1453-1474(1997).
[3]
RP SEQUENCE OF 307-826 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251358; PubMed=9097040;
RA Itoh T., Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Kujala H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,
RA Mizobuchi K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y.,
RA Nishimoto H., Nishio Y., Oshima T., Saito N., Sampei G., Seki Y.,
RA Sivasubram S., Tagami H., Takeda J., Takemoto K., Wada C.,
RA Yamamoto Y., Horiuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 40.1-50.0 min region on the linkage map."
RL DNA Res. 3:379-392(1996).
CC -!- FUNCTION: INVOLVED IN THE EXPORT AND ASSEMBLY OF A FIMBRIAL
SUBUNIT ACROSS THE OUTER MEMBRANE (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane
(BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U00007; AAA60473.1; -
DR EMBL; AE000300; AAC75170.1; -
DR EMBL; D90848; BAA15975.1; -
DR EcoGene; EGI1988; yehB.
DR InterPro; IPR000015; Fimb_usher.
DR Pfam; PF00577; Usher; 1.
DR PROSITE; PS01151; FIMBRIAL_USHER; 1.
KW Hypothetical protein; Outer membrane; Transmembrane; Fimbria;
KW Transport; Signal; Complete proteome.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 826 HYPOTHETICAL OUTER MEMBRANE USHER PROTEIN
FT YEHF.
FT DISULFID 809 825 POTENTIAL.
SQ SEQUENCE 826 AA; 92282 MW; 1FA541B1A21675CB CRC64;

Query Match 3.5%; Score 146; DB 1; Length 826;
Best Local Similarity 20.5%; Pred. No. 0.33;

```

```

Matches 148; Conservative 92; Mismatches 306; Indels 176; Gaps 34;
QY 135 ATVMQIETE-ITNOYISQYNT-----EITVK-----OTMLDGNRVKL----DMTFAEG 179
D 135 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 42 ANIRLDNDQPLPGQYDIDYVKNQWGRGYEIVKNDPQETCLSRVIRKLGINSDFASG 101
QY 180 KPARVVDINIIQNOHFSADLIDVLAIKNDKINPLSKADRYTQKLVTSLENLAKYLNA 239
D 180 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 102 KQCLTFQOLVGGSYTWDIGVF-----RLDFSVPAQWVEEL 139
QY 240 GFVFEIKDAKLINEDKNRIFVEISLHEGQYRFQTOFLGNLTYYTQAELEALLKFAE 299
D 240 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 140 GYVPE-----NWERGINAFYTSYL--SQY-YSDYKASGNNKSYVVRNSGLNLLGW 189
QY 300 EGFQAMLEQTTNN-----ISTKEGDDGYAAQIRPVTRINDESRITVDVEYIDPVHPVY 354
D 300 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 190 QLHSDASFSKTNNPNPVGWKSNLTLYLGRG--FAQLGLTLRVGDMYTSSDI--FDSVFRG 244
QY 355 VRRINFTGNFKTQDEVLRREMRQL-----EGALASNOKIQLSRARLMTGFEFKH 403
D 355 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 245 VR-----LFRDMQMLPNSKQNFPRVQGLAOSNALVTIE-----ONGF--- 282
QY 404 TVDTRVPNSPDQV-----DVNVVEQPSGSTIAAGYSQSGVTFQFQDSQN 453
D 404 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 283 -VYQKEVPPGPFPAITDLQLAGGADLDVSVKEADGVSVTYLVFYAAVNN-LQPGVSKY 340
QY 454 NEMGTGKHVNASFSRSE-TREVYSLGNTNPFYTVNGVYSQSLSGYRYKTKYDNKNISVYL 512
D 454 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 341 DLAAGRSHIEGASKQSDFOAGYQYGFNN-LLTYLGGSMVANNYYAFT----- 387
QY 513 DSYGGSLSYGPIDENQIRISFGLNADNTKLHGGFMCISNVKOLMADGKIQVDNNGIPD 572
D 513 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 388 -----LGAGW---NTRIG-ALSVDATKSHSKQDNGD-----VFDGQSYAIYN--K 427
QY 573 FKHDYTYTNAIIGWNTSSLDPRVPFPGQMSHSDLVFVGFGDKTHQKVY----- 621
D 573 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 428 FVSQTSRFGLAARYSRSDRYTFNDHVWANNKD-----NYRRDENDVYDIADYQNDG 482
QY 622 QGNIRPFFIKKSVLRGKVLGYGNLPPFYENFVAGVSGVGRGVDQS-SLGPSSQAV-LTA 679
D 622 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 483 RKNSFSNNMSQSLPEGWGSV---SLSTLRDYGWGRSGSKDYOLSYNNLRISYTLAA 538
QY 680 RRCQQTTLGVEVCGNALATFGSELILPLPFKGDWIDQV---RPVIFEGGVQVDTTGMOK 736
D 680 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 539 SQAYDENHHHEKRFN-----IFISIPF--DWGDDVSTPRQLYMSNSTTDFDQGFAS 588
QY 737 QTIDLTFQKDPQATACONAKAANRPILLTQDKQLRYSAGVCAWTYPIGLPISIAKPLNK 796
D 737 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 589 NN---TGLSGTVGRDQFNYGVN--LSHOGQNETTAGANLNTWNPATVNGSYSSSTY 643
QY 797 KQ 798
D 797 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 644 RQ 645

RESULT 15
PIP_LACLC
ID PIP_LACLC STANDARD; PRT; 1902 AA.
AC P16271;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE PI-type proteinase precursor (SC 3.4.21.-) (Wall-associated serine
proteinase).
DE PRTP.
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OG Plasmid pWV05.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WG2;

```









Db 2 KRELLPAL-LSALMIAEVHABS--FTVSDIRVNGLQVSAGSVFAALPLNVGETIDDOAL 58  
Qy 67 ADGVKALVATGNFSDVQVYHOEGRIIYQVTERPLIAEINFEGNRLIPKEGLQGLKNAGL 126  
Db 59 VQATRSFLKTFGFFQIOLGRDGNVLVTVVERPSTSSIEIEGNAKISKEDLLKGLKQSGL 118  
Qy 127 AVGOPLKQATQVMTETELTNOYISQGYNTEITVKOTMLDGNRVKLDMTFAEGKPARVVD 186  
Db 119 AGEIFQATLEGVNEQLQRYVAGQRYSAEINAEVIPPQNRVALKININEGTVAASH 178  
Qy 187 INIIGNOHFSDADLIDVLAIK-DNKINPLSKADRYTOEKLVTSLNLRKAYLNAGFVFE 245  
Db 179 INVGNIVFSEEDLTLFELKTTNWSFFKNDKAYAREKLSGDLERLSYLDRGYNMD 238  
Qy 246 IKDAKLINEDKNRIFVEISLHEGEQYRFGQTQFLNLTYYQAELEALLKFAKEGFSQA 305  
Db 239 IASTQVSIITPKKVIYITVINEGEKYTIRDKLTGDLKVPBEEVKRLLLVQKGOVESRK 298  
Qy 306 MLEOTTNNISPKFGDGYVYVIAQIRPVTRINDESRVDVEYIDVPVHPVYVRINFTNFK 365  
Db 299 VMTTSLITRLNGEYITFANVGVPEAHDDDKTVSVTFVVDPGKRAYVNRINFRGNTK 358  
Qy 366 TQDEVLRREMRLQEGALASNOIKIOLSRARLMRTGFFKHVTVTRPVVPSDQVDVNEFWE 425  
Db 359 TEDEVLRREMRLQEGWASTYLDIOSKARLERLGYFKEVNVETPAVGTDDQVDVNSVE 418  
Qy 426 EOPSGSSTIAAGYSQSGVTFQFDVSONNFMGTGKHVNASFSRSETREVYSLGMTNPFYT 485  
Db 419 EOPSGSITASVGFASAGLILGGSISONNFLTGTGNKVYSIGLTRSEYQTRYNFGVDPYT 478  
Qy 486 VNGVQSLSGYRKTGYD--NKNISNVLDYSGSLSGYGPIDENQRIISFGLNADNTKLH 543  
Db 479 VDGSLGYNFAYKTDDELVDVASTVNSLGAWSIGYPISTSLTYGLSVQRDQID 538  
Qy 544 GGRPMGINSYKQLMADGGKIQVDNNGIPDKHDYTYTNNAILGNWYSLSLDRPVFPPTQGMH 603  
Db 539 TGRYT-VDEIYDFLDKEG---DN-----FTNFKASIGWSESTLNGKVLATRGHSQ 584  
Qy 604 SVDL--TVGFGDKTHOKVYVYOGNIYRPIKKSVLRGYAKLGYGN-----NLFFYENFYA 655  
Db 585 SLTLETLPGSDLSFYKIDYRGQVFAPLDNTYMRFTHELGYGDGYGSGTERLFPYENYA 644  
Qy 656 GGXGSRVGYDQSSILGPRSQAYLTARRGQQTTL-----GEVVGGNALAT 698  
Db 645 GGNSVRGFKDSTLGPSTFSV-ARNPDGTPMKNOGPDCKRYTDPDQDPEAFEGGNILIT 703  
Qy 699 FGSELILPLFPKGDWIDQVPVPIEGGVQVFDTTGMDKQTDIDLTQFKDPOATAEQNAKAA 758  
Db 704 GGAELLPLFPVKDQ-RQLRTVLFDVGVSTFDTDCPTKTTNCDGK-----749  
Qy 759 NRPLLTQDKOLRYASAGVATWYPIGPLSISYAKPLNKKQNDQTDVQFOIGSVF 813  
Db 750 -----TDNLASVGVGLTWITLALGPLSFLATPIKKPDNAETQVFPFSLGQTF 797

## RESULT 2

Q9S341 PRELIMINARY; PRT; 797 AA.  
AC Q9S341;  
DT 01-MAY-2000 (TRENBLrel. 13, Created)  
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)  
DE OUTER MEMBRANE ANTIGEN.  
GN OMA.  
OS Photorhabdus luminescens (Xenorhabdus luminescens).  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Photorhabdus.  
OX NCBI\_TaxID=29488;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HM.  
RA Chatonnet-Marton P.I., Givaudan A., Ianois A., Boemare N.E.;  
RT "Photorhabdus luminescens genomic region homologous to 4.0 minute

RT Escherichia coli region promotes pleiotropic phenotypes.\*;  
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ236920; CAP51929.1; -  
DR InterPro: IPR000184; Bac\_surfaG\_D15.  
DR Pfam: PF01103; Bac\_surfaceAg; 1.  
SQ SEQUENCE 797 AA; 88713 MW; 39E72E7ABD9C14F3 CRC64;

Query Match 29.8%; Score 1254; DB 2; Length 797;  
Best Local Similarity 34.68; Pred. No. 7e-61;

Matches 282; Conservative 161; Mismatches 334; Indels 38; Gaps 12;

Qy 14 MTAVMVMVSTHAQAAD-FMANDITITGLQRTVIESQSVLFFRLGQVSVSNQLADGVKA 72  
Db 6 LLIASLLFGSAAAYGADGVFVQDIHFEGLQRVAVGAALLNMPVRGVDVSDIGRIHA 65  
Qy 73 LYATGNSDVQVYHOEGRIIYQVTERPLIAEINFEGNRLIPKEGLQGLKNAGLAVGQPL 132  
Db 66 LFATGNFEDVRVLRDGNNTLIVQKERPTIASITFSGNKSVKDDMLKQNLASHVRVGEAL 125  
Qy 133 KOATVQMIETELTNOYISQGYNTEITVKOTMLDGNRVKLDMTFAEGKPARVVDINIEN 192  
Db 126 DRTMISNIERGLEDFYISVGKYNASVKNVTPLPNRVLDLKLVEAGVSAKIQIINIVGN 185  
Qy 193 QHFSADADLIDVLAIKDNKINP---LSKADRYTOEKLVTSLNLRKAYLNAGFVFEIKDA 249  
Db 186 KSFSSDELLNRFQLRDDV--PWNLTADQYKQKLGDLLEALRSFYLDRGYARENDIST 243  
Qy 250 KLNINEDKNRIFVEISLHEGEQYRFGQTQFLNLTYYQAELEALLKFAKEGFSQAMLEQ 309  
Db 244 QVSLTPDKKGIYVITNTEGDOYKISGIDLNGNAGYOSEITKLAIEPGLSYNGTQVTK 303  
Qy 310 TTNISTKFGDDGYVYVIAQIRPVTRINDESRVDVEYIDVPVHPVYVRINFTNFKTODE 369  
Db 304 MENDIKNLLGRYGVAYPRVMTQPEINDQDKTVKLHVINDAGNRYVRKIRFSGNDTKDS 363  
Qy 370 VLREMRQLEGALASNOIKIOLSRARLMRTGFFKHVTVTRPVVPSDQVDVNEFVEOPS 429  
Db 364 VLREMRQMERAWLGLSDLVELGKERLNLGYFETVDVETQRIPGSPQVDVYVYKVRNT 423  
Qy 430 GSSTIAAGYSQSGVTFQFDVSONNFMGTGKHVNASFSRSETREVYSLGMTNPFVYNGV 489  
Db 424 GSLNFGVGFCTESGVSFQIGAQDNLGTCNAGINASKNDYSTYAELSFTDPTFTINGV 483  
Qy 490 QSLSGYRYKTKYDNKNISNVLDYSGSLSGYGPIDENQRIISFGLNADNTKLHGGRFMG 549  
Db 484 SLGGRVYINDFRADDAELSGYTNQSYGSLGFLPPIINNENSLNFGNLYIHNSL-SDMLPQ 542  
Qy 550 ISNVKOLMADGGKIQVDNNGIPDKHDYTYTNNAILGNWYSLSLDRPVFPPTQGMHSDV-L 607  
Db 543 VAMVYLRSMGKPDLESKA--EFKAD--DFALTMGWTYNNLDRGFFPTSGVSKSLNGKV 598  
Qy 608 TVGFGDKTHOKVYVYOGNIYRPF--IKKSVLRGYAKLGYGN-----NLFFYENFYAGYGS 660  
Db 599 TIPGSDNEFYKVTLDTSAYYPIINDRTWVILGRSRLYGDLGGLGKELPFYENFYAGSSST 658  
Qy 661 VRGYDOSSLGPRS-QAYLTARRGQQTTLGVEVVGGNALATFSELILPLPF-KGDWIDQVR 718  
Db 659 VRGFRSNNIGPKAIYLYKDGSPKSPSRDAVGNAMAVASLELITPTPLPDKYSNSVR 718  
Qy 719 PVTFIEGGVQVFDTTGMDKQTDIDLTQFKDPOATAEQNAKANRPLLTQDKOLRYASAGVAT 778  
Db 719 TSFFIDSGVTWMDTWNDSAVM-----KSGIPDYSKPGNIRVSAGIALQ 762

## RESULT 3

Q9KPW0  
ID Q9KPW0  
AC Q9KPW0;  
PRELIMINARY; PRT; 803 AA.

```
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SURFACE ANTIGEN.
GN VC252.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OC NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Olin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483(2000).
DR EMBL: AE004297; AAF95396.1; -.
DR TIGR: VC252.
DR InterPro: IPR000184; Bac_surfaAg_d15.
DR Pfam: PF01103; Bac_surface_Ag; 1.
DR Complete proteome.
KW SEQUENCE 803 AA; 89930 MW; 1998B3838F6C041D CRC64;

Query Match 28.6%; Score 1203; DB 16; Length 803;
Best Local Similarity 33.4%; Pred. No. 4.5e-58;
Matches 278; Conservative 155; Mismatches 328; Indels 72; Gaps 15;

Qy 16 MAVVMVMSHQAQAD-FNADITITGLQVYIESLQSVLPRLQGVSVENLADGVKALY 74
Db 8 LATLLTSVANGAEKFFVQDIQGLQVVALGAALLKMPVRVGSVDSDVAVIKALY 67
Qy 75 ATGNFSDVQVYHQEGRIIYQTERPLIAEINFEGNRLIPKCEQLQGLKAGLAVGQPLQ 134
Db 68 SSGNFEDVKVLRDGNTLAVQYKERTIASVSFSGNKAKEQLKONLEASSIRVGEALDR 127
Qy 135 ATVQMIEFTELNYIQSYQYTFEITVTKOTMLDGNRVKIDMTFAEGKPARVVDINIIGNOH 194
Db 128 TTLSNIEKGLEDFYYSVGKNATKAVVTPLPNRADLKVFETEGSAKIQOINFIQNV 187
Qy 195 FSDADLDVLAIKNKI--NPLSRADRYTOEKLVTSLNLRKYLINAGVFPEIKDAKLN 252
Db 188 FSDELLSRFNLVDVAMWNLAD-DKYKQVLGADIEALRTYYLDRLGKLFQVDSTQVA 246
Qy 253 INEDKNRIFVEISLHEGQYRFQGTQFLGNLTYYTQAELEALLKFAEFGFSQAMLEQTTN 312
Db 247 LSPDKGGVITLNLNEGEPYTVSKVQFRLGMLGKEAETSLIPFELGYNGSAVTRLEE 306
Qy 313 NISTKFGDDGYIAQIRPVTRINDESRVDVEYIDPVHPVVRINTGKFNKQDEVLRL 372
Db 307 SVKVLGSEGYAYFQVRIPEFDDEKQGVSLVHVHVEAGKRVVVRDIREVGNNSRDEVLRL 366
Qy 373 REMRQLEGALASNOKIOLSRRLMTGFKHVTVDTRVPNSPDQVDVNFVVEEQSGSS 432
Db 367 REMRQMGESWLNSKDIETGKTRLRNLGFFETVEQTVRVPSGSDVDLVYSVKEANSNV 426
Qy 433 TIAAGYSQGGVTFQFQVSONNFMTGKHVNASFSSRETREYVSLGNTNPFYFTVNGVSQS 492
Db 427 NFGVGYGTESGVSFQVGLQDNFLGSGNRVGVNMINDYQKNLTLEYRDPYWNLDGVSIG 486
Qy 493 LSGYRKTKYDNKNISNVLDYSGLSYGYPIDENQRISFGLNADNTKLHGRPMGTSN 552
Db 487 GKVEYNQFEASEAGIVDYTNESYGTSLTWGFPFDLNFREFGIGYTHNKI--GNLTPYLQ 544
Qy 553 VKQLMA-----DGGKIQVDNNGIPDPFKHDYTYTNAILGNWYSSLDPRVPTQMSHS- 604
Db 545 VENFLAAQASIDSGGNLLTDD-----FDINLSWTRNRLNNSYFFTAG-NHQR 591
```

```
Qy 605 --VDLTVFGDKTHOKVYQGNIRPFIKKS-----VLRGYAKLGYGNN-----LP 648
Db 592 AFYKMTVPGSDAQYFKLOYDVQYFPLTKKHEFTLLRG--RLCYNGYQOTGKDNLP 649
Qy 649 FYENFYAGYGVSRGYDQSSLSGPRS--QAYLTARRGQQTTLGVEVGNALATFGSELILP 706
Db 650 FYENFYAGGTSLRGFGSNSAGPKAVYRDYSNGNSGSDTATDDSVGNAIALASVELIVP 709
Qy 707 LPFKGDWI--DOVRPVIFIEGGQVFDITGMDKQITD-----LTQFKDPQATAEQNAKANR 760
Db 710 TPFASEEARNQRTSIFYDMA SVMDTEFDYRGKADYGNQYVYDSDP----- 756
Qy 761 PLLTDQKQRLYSAGVAGATWTPIGPLSISYAKPLNKKQNDOTDTVOFOIGSVF 813
Db 757 -----TNVRSYSGVALQWSPMGPLVFLSLAKPKIKYEGDDEFFTFITGRTF 803

RESULT 4
Q9R2E3 PRELIMINARY; PRT; 785 AA.
AC Q9R2E3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 87.9 KDA PROTEIN.
GN YZZN.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA Takemoto K., Mori H., Murayama N., Kataoka K., Yano M., Itoh T.,
RA Yanamoto Y., Inokuchi H., Miki T., Hatada E., Fukuda R., Ichihara S.,
RA Mizuno T., Makino K., Nakata A., Yura T., Sampei G., Mizobuchi K.;
RT "Systematic sequencing of the Escherichia coli genome: Analysis of the
RT 4.0 - 6.0 min (189,987 - 281,416bp) region.";
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=94261430; PubMed=8202364;
RA Fujita N., Mori H., Yura T., Ishihama A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of the
RT 2.4-4.1 min (110,917-193,643 bp) region.";
RL Nucleic Acids Res. 22:1637-1639(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=82059454; PubMed=6272196;
RA An G., Bendiak L., Mametlak A., Friesen J.;
RT "Organization and nucleotide sequence of a new ribosomal operon in
RT Escherichia coli containing the genes for ribosomal protein S2 and
RT elongation factor Ts.";
RL Nucleic Acids Res. 9:4163-4172(1981).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=93077430; PubMed=1447125;
RA Yamanaka K., Ogura T., Niki H., Hiraga S.;
RT "Identification and characterization of the smba gene, a suppressor of
RT the mukB null mutant of Escherichia coli.";
RL J. Bacteriol. 174:7517-7526(1992).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA Smallshaw J., Kelln R.;
RT "Cloning, nucleotide sequence and expression of the Escherichia coli
RT K-12 pyrH gene encoding UMP kinase.";
RL Genetics 11:59-65(1992).
RN [6]
```

RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=94240115; PubMed=8183897;  
 RA Janosi L., Shimizu I., Kaji A.;  
 RT "Ribosome recycling factor (ribosome releasing factor) is essential  
 RL for bacterial growth";  
 RN Proc. Natl. Acad. Sci. U.S.A. 91:4249-4253(1994).  
 [7]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=91317739; PubMed=1860827;  
 RA Shimizu I., Kaji A.;  
 RT "Identification of the promoter region of the ribosome-releasing  
 RL factor cistron (frr)";  
 RN J. Bacteriol. 173:5181-5187(1991).  
 [8]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=90062117; PubMed=2684966;  
 RA Ichikawa S., Kaji A.;  
 RT "Molecular cloning and expression of ribosome releasing factor.";  
 RL J. Biol. Chem. 264:20054-20059(1989).  
 [9]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=86008268; PubMed=2995358;  
 RA Icho T., Sparrow C.P., Raetz C.R.H.;  
 RT "Molecular cloning and sequencing of the gene for CDP-diglyceride  
 RL synthetase of Escherichia coli";  
 RN J. Biol. Chem. 260:12078-12083(1985).  
 [10]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=91123198; PubMed=1991717;  
 RA Hirvas L., Koski P., Vaara M.;  
 RT "The ompH gene of Yersinia enterocolitica: cloning, sequencing,  
 RL expression, and comparison with known enterobacterial ompH  
 sequences";  
 RN J. Bacteriol. 173:1223-1229(1991).  
 [11]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=91100302; PubMed=1987124;  
 RA Dicker I., Seetharam S.;  
 RT "Cloning and nucleotide sequence of the fira gene and the fira200(ts)  
 RL allele from Escherichia coli";  
 RN J. Bacteriol. 173:334-344(1991).  
 [12]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=88058790; PubMed=2824445;  
 RA Crowell D., Reznikoff W., Raetz C.;  
 RT "Nucleotide sequence of the Escherichia coli gene for lipid A  
 RL disaccharide synthase";  
 RN J. Bacteriol. 169:5727-5734(1987).  
 [13]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=88058791; PubMed=3316192;  
 RA Tomaszewicz H.G., McHenry C.S.;  
 RT "Sequence analysis of the Escherichia coli dnaE gene.";  
 RL J. Bacteriol. 169:5735-5744(1987).  
 [14]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=88139188; PubMed=3277952;  
 RA Coleman J., Raetz C.;  
 RT "First committed step of lipid A biosynthesis in Escherichia coli:  
 RL sequence of the lpxA gene";  
 RN J. Bacteriol. 170:1268-1274(1988).  
 [15]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;

RX MEDLINE=92380982; PubMed=1355089;  
 RA Li S., Cronan J.;  
 RT "The genes encoding the two carboxyltransferase subunits of  
 RL Escherichia coli acetyl-CoA carboxylase";  
 RN J. Biol. Chem. 267:16841-16847(1992).  
 [16]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=90094229; PubMed=1688424;  
 RA Zhou Z., Syvanen M.;  
 RT "Identification and sequence of the drpA gene from Escherichia coli.";  
 RL J. Bacteriol. 172:281-286(1990).  
 [17]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=90370122; PubMed=2203971;  
 RA Eriani G., Delarue M., Poch O., Gangloff J., Moras D.;  
 RT "Partition of tRNA synthetases into two classes based on mutually  
 RL exclusive sets of sequence motifs";  
 RN Nature 347:203-206(1990).  
 [18]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=93094132; PubMed=1459951;  
 RA Gervais F.G., Drapeau G.;  
 RT "Identification, cloning, and characterization of rcsF, a new  
 RL regulator gene for exopolysaccharide synthesis that suppresses the  
 RN division mutation fts284 in Escherichia coli K-12";  
 RP J. Bacteriol. 174:8016-8022(1992).  
 [20]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=94124004; PubMed=7904973;  
 RA Allikmets R., Gerrard B., Court D., Dean M.;  
 RT "Cloning and organization of the abc and mdl genes of Escherichia  
 RL coli: relationship to eukaryotic multidrug resistance";  
 RN Gene 136:231-236(1993).  
 [21]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=93011013; PubMed=1396599;  
 RA Condon C., Phillips J., Fu Z., Squires C., Squires C.;  
 RT "Comparison of the expression of the seven ribosomal RNA operons in  
 RL Escherichia coli";  
 RN EMBO J. 11:4175-4185(1992).  
 [22]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=79012454; PubMed=358189;  
 RA Young R.A., Steitz J.A.;  
 RT "Complementary sequences 1700 nucleotides apart form a ribonuclease  
 RL III cleavage site in Escherichia coli ribosomal precursor RNA";  
 RN Proc. Natl. Acad. Sci. U.S.A. 75:3593-3597(1978).  
 [24]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=93116053; PubMed=1474579;  
 RA Cormack R., Mackie G.;  
 RT "Structural requirements for the processing of Escherichia coli 5 S  
 RL ribosomal RNA by RNase E in vitro";

Query Match	28.5%	Score 1197	DB 2	Length 785
Best Local Similarity	34.4%	Pred. No. 9.3e-56		
Matches	281	Conservative 145	Mismatches 319	Indels 72
				Gaps 14
QY	35	DITTGILQRTVIESLSVLPRLGOVSVENOLADGVKALYATATGNFSDVQVYHQBGRRIYQ	94	
DB	3	DIHFEGLQRTAVGAALLSMPVRTGDTVNDIEDISNTRALFATGNFDRVLRDGTLLVQ	62	
QY	95	VTERPLTAETNFEGNRLIPREGLOGLKNAGLAVGQPLKQATVOMIETELTNQISQGY	154	
DB	63	VKERPTASITFSNGKSVKDDMLKNLEASGRVRGESLDRTTIADIEKGLDEDFYISVGKY	122	
QY	155	NTERITVQTMDLGNRVRKIDMTFAEGKPARVVDINIIGNHFSADLLIDVLAIKD-----N	209	
DB	123	SASVAKAVTLPPLRRVRDLKLVFQEGSVNAEQIINVGNHAFITDELISHFOLRDEVPWN	182	
QY	210	KINPLSKADR-YTQEKLVTSLENIKAKYLNAGFYRFEIKAKLINEDKNRIFVEISLHE	268	
DB	183	VW-----GDRKYQKLAGDLLETURSYLYDRGYARFNIDTSQVSLTPDKGIYTVNITE	237	
QY	269	GEQYRFGQTQFLGNLTQYAELEALLAKFAEEGFSQAMLEQTNNISKTKGDDGYAAQI	328	
DB	238	GDQYKLSGEVSGNLGASHAEIQLTKEPGLYNGTKVTKMEDDIIKLLGRYGYAPRV	297	
QY	329	RPVTRINDESRTVDVEYIDPVHVYVRRINFTGNFTQDEVLRRERMOLEGASLAKQKI	388	
DB	298	QSMPEINDADTKVLRYNVNADGNFYRKIRFEGNDTSKDAVLRERMQEMAGLGSGLV	357	
QY	389	QLSRAIRLMTGFFKHVITDTRVPVNSPDQVDVNFVVEEPQSGSSTIAAGYSQSGVTFQF	448	
DB	358	DQGERLNRLLGFFETVDTDTQVRPSPDQVDVYKVKERNTGSFNGIGYGTESGVSFQA	417	
QY	449	DVSONNFMGTGKHYNASFSRSETREYVSLGNTNPFYFVNGVYSQSLGYRKYTKYDNKNIS	508	
DB	418	GVQDDNMLGTGYAVGINGTKNDYQTYAELSVTNPFYFVDGYSGLGRFLYNDFOADDADLS	477	
QY	509	NYVLDSYGGSLSYGYPIDENORISFGLNADNTKLHGGRFMGISNVKOLMA-----DGGK	562	
DB	478	DYTNKSGTVDTLGFPINEYNSLAGLGYHNSL-----SNQOQVAMWRVLYSMGE	529	
QY	563	IQVDNIGIPDKHDYTYNAILGNWYSLSLRPVPFTQG--MSHSVDLTVFGDKTHQKV	620	
DB	530	HPSTSDQNSFKTDDETFN--YGWYTNKLDRGYFPTDGSRVNLTKGVTIPGSDNEYKYVT	587	
QY	621	YQGHNIYRPF--IKSVLRGYAKLGYGNL-----PFTENFYAGYGVSRGVDQSSLPGRS	673	
DB	588	LDTATYVIPIDDDHKWVYLGRTRWGYGDLGKEMPFYENFYAGGSSTVRFSQSTIGPKA	647	
QY	674	-----QA-----YLTAARGQQTTL--GEVYGGNALATFGSELILPLPKGD-WIDQ	716	
DB	648	VYFPHQASNYDPDYDECATQDGAKDLCKSDDAVGGNMAVASLEFPTPTFISDKYANS	707	
QY	717	VRPFIIEGGQVFTTGMCKQTIDLTQFKDQATAEQNAKAANRPLLTQDKQLRYSAGV	776	
DB	708	VRTSFFDMGTVDNTNDWSSQYSGYPDYSDP-----GNIRMSAGIA	748	
QY	777	ATWYTPIGPLISYAKPLNKKQNDQDTVQFQIGSVF	813	
DB	749	LQWMSPLGLPLFSYAQPFQYDGDKAQEQFQFNIGKTW	785	
RESULT	5			
Q9PEI2		PRELIMINARY;	PRG	784 AA.
ID	Q9PEI2			
AC	Q9PEI2;			
DT	01-OCT-2000 (Tremblrel. 15, Created)			
DT	01-OCT-2000 (Tremblrel. 15, Last sequence update)			
DT	01-DEC-2001 (Tremblrel. 19, Last annotation update)			
DE	OUTER MEMBRANE ANTIGEN.			
GN	XF1046.			
OS	Xylella fastidiosa.			

```
QY 432 STIAAGYSQGGVTFQFDVSONNFMGTGKHVNASFSSRSETREVYSLGTMTPYFTVNGVSQ 491
Db 432 FOYGLGYSKYTYTTSVOLSONFLSGNRVSDASKRSYQDRYSFSYTNPFDTNGVSL 491
QY 492 SLGGYRKTKYDNKNISNYVLDYSGGSLSYGYPIDENQRIISFGLNADNTKLHGGREMGIS 551
Db 492 GYNLAYOKLDYSDFNAAQYSKRMSGQTFIGIPTENDTYSWVIGADSNQI--TTFPG-S 548
QY 552 NVQLMADGCKIQVDNNGIDPFKHDTTYNAIILGWNYSSSLDRPFFPQGHSHSDVLTGVF 611
Db 549 TPRAIID-----YIDAVG-----QRTFRATELGGWARDTRNDYFMPNLGMRYRIGAEVTL 599
QY 612 GDKT--HOKVVYOGNIYRPPIKKSVLGRYAKLGYGNN-----LPGY 650
Db 600 PGSTIKYKYNQISKWPIIPALVNLTRLEVGYGDDYKSHTRILPDGTIVATASGLPFF 659
QY 651 ENFYAGYGSVGRVYDQSSSLGPRSOAYLTARRGOOTTLGEVVGNGALATFSGELILPLPFFK 710
Db 660 ENFYAGTNSVRGFRDNTLGRSE--VTALYNQ---GQPLGGSEKTVGSTEMYFPKLFED 713
QY 711 GDWIDQVRPIFIEGGGVFTTGMCKQTDIDLTQPKDPAQAEONAKAANRPLLTQDKQLR 770
Db 714 S---PSARISAFDLDFGVNF-----GVNNFKA-----NELR 741
QY 771 YSAGVGATWTPIGPLSISYAKPLNKKONQDQTDVQFQIGSVF 813
Db 742 ASSGVALLWRAPIGPISISYAFPIKKENDEIERLQTFEGQGF 784
RESULT 6
Q9XJ10 PRELIMINARY; PRT; 797 AA.
AC Q9XJ10:
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, last sequence update)
DE OUTER MEMBRANE PROTEIN OMP85.
GN NMB0182.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
[1]
SEQUENCE FROM N.A.
RP STRAIN=MC58 / SEROGROUP B;
RC MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., Desoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Citterone H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Masiagnani V., Pizzo M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
MC58."
RL Science 287:1809-1815(2000).
DR EMBL; AE002375; AAF40639.1; -.
DR TIGR; NMB0182; -.
DR InterPro: IPR000184; Bac_surfag_D15.
DR Pfam; PF01103; Bac_surfaceA9; 1.
KW Complete proteome.
SQ SEQUENCE 797 AA; 88436 MW; B68BFC5A43D22EB8 CRC64;
```

```
Query Match 26.58; Score 1115; DB 16; Length 797;
Best Local Similarity 32.58; Pred. No. 3.le-53;
Matches 266; Conservative 154; Mismatches 361; Indels 38; Gaps 14;
QY 10 QVSAMTAVMMVMVMTAAQADFMANDITITGLQRTVIESLQSVLFFRLGQVVSQNLADG 69
Db 2 KLKQIASALMLMGISPLALADFTIQDIRVEGLQRTPESTVFNLPVKVGDYNDTHGSAI 61
```

```
QY 70 VKALYATGNFSDVOYVHOBGRIIYQVTERPLIAINEFGRNLIPKEGLQEGLNAGLAVG 129
Db 62 IKSLYATGTFDDVRVETADGOLLTVTERPTIGSLNITGAKMLQNDAIKKLSEFGLAQS 121
QY 130 OPLKQATVQMLELTNTQYISOGYYNTEITVKQTMLDGNRVKLDMTFAEGKPARVDINI 189
Db 122 QYFNOATLNOAVAGLKBEGYLGKGLNIQITPKVKTLARNRVDDIITIDEGSAKITDEF 181
QY 190 IGNOHFSDADLIDVLAJDKNKI-NPLSKADRYTOEKLVTSLENRAKYINAGFVRFEIKD 248
Db 182 EGNQVYSDRKLMRQMSITTEGIWTLTRESNOFNEOKFAODMEKYTDFQNNGYDFRILD 241
QY 249 AKLNINEDKNRIFVEISLHEGEYRFQGTQFLGNLT-YTQAEALALKFAEEGFSQAML 307
Db 242 TDIQTNEKTKQITKITVHEGGERGWKVSIEGTDNEVPKAELEKLLTMTKPKGKVEROOM 301
QY 308 EOTTNNISTKFGDDGYAAQIRPVTRINDESRTVDVEYYIDPVHPVYVRRINFTGNKFTQ 367
Db 302 TAVLGEIONRMGSAGYAYSEISVQPLPNAETKTVDVLIHEGGRKIYYNEIHTGNKTR 361
QY 368 DEVLRRERMLQLEGALASNQIKQLSRARLMRTGFFKHVTVDTRPVNPSQDQDVNFVVEEQ 427
Db 362 DEVYRELRLQMESAPYDTSKLQRSKERVLLGYFDNVQFQDAPLAGTDPKVDLNMSLTER 421
QY 428 PGSSTIAAGYSQGGVTFQFDVSONNFMGTGKHVNASFSSRSETREVYSLGTMTPYFTVN 487
Db 422 STGSLDLSAGWVQDTGLVMSAGVSQDNLFGTGKSAALRASRSKTTLNGSLSTDFPYFTAD 481
QY 488 GVSQSLSGYR--KTKYDNKNISNVLDYSGGSLSYGYPIDENQRIISFGLNADNTKLHGG 545
Db 482 GVSIGVDYGRAPDRKASTSIKQYKTTAGAGIRMSVPVTEYDRVNEGLVAEHLTVN-- 539
QY 546 RFMGISNVKQLMADGGKIQVDNNGIPDPFKHDYTYTNAIILGWNYSSSLDRPFPPTQGMHSV 605
Db 540 --TYNKAPKHVADFIKYKGTG--TDGSFKGWLYKGVGWGRNKTDSALWPTRCYLTGV 595
QY 606 DLTGVF-GDK--THQKVYQGYNIYRPFIKKSVLK-----GYAKLGV--NNLPFFYENFVAG 656
Db 596 NAEIALFGSKLQYYSATHNQTWTFPLSKTFTLMLGGEVGIAG-GYGRTKETIPFFENFYGG 654
QY 657 GYGSVRGVYDQSSSLGPRSOAYLTARRGOOTTLGEVV--GGNALATFSGELILPLPKGDWI 714
Db 655 GLGSVRGYESTLGPK-----VYDEYGEKISYGGNKKANYSAEILLPMPGAKD-A 703
QY 715 DQVRPVIFIEGGQVFDTTGMDKQTDIDLTQPKDPAQAEONAKAANRPLLTQDKQLRYAG 774
Db 704 RTVRLSLFADAGSVMDGKTYDDNSSAT-----GGRVQNIYCGAGNTHKSTFTTNELRYSAG 758
QY 775 VGATWYTPIGPLSISYAKPLNKKONQDQTDVQFQIGSVF 813
Db 759 GAVTWLSPLGPMKFSYAYPLKKKKPEDEIQRFQFQJGTTF 797
RESULT 7
Q9XJ31 PRELIMINARY; PRT; 797 AA.
AC Q9XJ31:
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE OUTER MEMBRANE PROTEIN OMP85.
GN OMP85 OR NMA0085.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
[1]
SEQUENCE FROM N.A.
RP STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4a;
RC MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
```



RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,  
RT Whitehead S., Spratt B.G., Barrell B.G.;  
RA "Complete DNA sequence of a serogroup A strain of *Neisseria*  
meningitidis 22491";  
RL Nature 404:502-506(2000).  
DR EMBL; AL162752; CAB83401.1; -;  
DR InterPro; IPR000184; Bac\_surfAg\_D15.  
DR Pfam; PF01103; Bac\_surface\_Ag; 1.  
KW Complete proteome.  
SQ SEQUENCE 797 AA; 88404 MW; 65DE47E00C9E1D1F CRC64;

Query Match 26.4%; Score 1111; DB 16; Length 797;

Best Local Similarity 32.4%; Pred. No. 5.1e-53;

Matches 265; Conservative 155; Mismatches 361; Indels 38; Gaps 14;

```
QY 10 QVSAMTMAVMVMSTHAQAADFMANDITITGLQRTVIESLQSVLPFRGLQGVVSENQADG 69
DB 2 KLKQIASALMWLGISPLALADFTIQDIRVEGLQRTPESTVFVYLPVKVGDYNDTHGSAI 61
QY 70 VKALYATGNFSDVQYVHQEGRIYQVTERPLIAEFNFGNRLIPKREGLOGLKNAGLAVG 129
DB 62 IKSLYATGFDDVRVETADGQLLTVERPTIGSLNITGAKMLQDAIKKNLESFGLAQS 121
QY 130 QPLKQATVOMIETELTNQYISQYYNTEITVKQTMLDGNRVKLDMTFAEGKPARVVDINI 189
DB 122 QYFNQATLNOAVAGLKEEYLGKGLNIQITPKVTKLARNRVDITIDEGSKAKITDIEF 181
QY 190 IGNOHFSADLDIVLAIKDKNI-NPLSKADRYTQEKLVTSLENLRAKYNLAGVFEIKD 248
DB 182 EGNQVYSRDKLRQMSLREGGIWTLTRSNQFNEQFAQDMKVTDFYQNNGYDFRILD 241
QY 249 AKLNINEDKNRIFRVEISLHGEQYRFGQTFGLNLT-YTQAELEALLKFAEGFSQAML 307
DB 242 TDIOQNEKTKOTIKITVHEGGRFWMGVKYSIEGDTNEVPKAELEKLLTMKPKWYERQOM 301
QY 368 DEVLRRERMOLEGASLNQIOLSRALMRGTFKHVTVDRPVPNSPDQVDVNFVEEQ 427
DB 362 DEWVRRELQOMESAPYDTSKLQSKRVERVLLGYFDNVQFQDAVPLAGTDPKVDLNMSLTER 421
QY 428 PGSSTIAAGYSQSGVTFQFDVSONNFMGTGKHYNASFSESTREVSLSGMTNPFTVN 487
DB 488 GVSQSLSGYR--KTKYDNKNISNVLDYSYGSLSGYYPIDENQRIISFGLNADNTKLHGG 545
QY 482 GVSGLGYDYKAFDPRKASTSIKQYKTTTAGAGIRMSVPVTEYDRVNFGLVAEHLTVN-- 539
QY 546 RFMGISNVKQLMADGGKIQVDNNGIPDFKHDTYTNAILGNWYSSLDPRVPFTQGMHSV 605
DB 540 ---TYNKAPKHVADFIKKYKTDG--TDGSFGKWLKYGTVGWRNKRTDSALMPTRGYLTGV 595
QY 606 DLTGVF-GDK--THQKVYVQGGNIYRPIKKSVLK-----GYAKLGYG--NNLPFFYENFYAG 656
DB 596 NAEIALPGSKLQYYSATHNQWTFPPLSKTFTLMLGGEVGIAG-GYGRTKKEIPFFENFYGG 654
QY 657 GYGSVRGVDQSSLSGPRSQAYLTARRGQQTTLGEVV--GGNALATGSESLIPLPKPGDWI 714
DB 655 GLGYSVRGESGTGLPK-----VYDEXGEKISYGNKNKANYSAELLFPMPGAKD-A 703
QY 715 DQVRPVIIEGGOVDTTGMKQIDLTQFKDPQATAEQNAKAANRPLLTQDKQLRYSG 774
DB 704 RTVRLSLFADAGSVWVGKTYDDNSSAT-----GGRVQNIYGAGNTHKSTFTNELRYSG 758
QY 775 VGATWYTPIGPLSISYAKPLANKQNDQTDTFQFQIGSVF 813
DB 759 GAVTWLSPLGPKMSYAYPLAKKKPDEIQRFQFOLGTTFF 797
```

RESULT 8  
O30912 PRELIMINARY; PRT; 797 AA.  
AC O30912;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE OUTER MEMBRANE PROTEIN OMP85.  
GN OMP85.  
OS *Neisseria meningitidis*.  
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; *Neisseria*.  
OX NCBI\_TaxID=487;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HH.  
RA Manning D.S., Reschke D.K., Judd R.C.;  
RT "Omp85 proteins of *Neisseria gonorrhoeae* and *Neisseria meningitidis*  
are similar to *Haemophilus influenzae* D-15-Ag and *Pasteurella*  
*multocida* Oma87";  
RL Microb. Pathog. 23:0-0(1998).  
DR EMBL; AF021245; AAC17599.1; -;  
DR InterPro; IPR000184; Bac\_surfAg\_D15.  
DR Pfam; PF01103; Bac\_surface\_Ag; 1.  
SQ SEQUENCE 797 AA; 88539 MW; CF911B5F70B999CF CRC64;

Query Match 26.4%; Score 1109; DB 2; Length 797;

Best Local Similarity 32.4%; Pred. No. 6.6e-53;

Matches 265; Conservative 154; Mismatches 362; Indels 38; Gaps 14;

```
QY 10 QVSAMTMAVMVMSTHAQAADFMANDITITGLQRTVIESLQSVLPFRGLQGVVSENQADG 69
DB 2 KLKQIASALMWLGISPLAFADFTIQDIRVEGLQRTPESTVFVYLPVKVGDYNDTHGSAI 61
QY 70 VKALYATGNFSDVQYVHQEGRIYQVTERPLIAEFNFGNRLIPKREGLOGLKNAGLAVG 129
DB 62 IKSLYATGFDDVRVETADGQLLTVERPTIGSLNITGAKMLQDAIKKNLESFGLAQS 121
QY 130 QPLKQATVOMIETELTNQYISQYYNTEITVKQTMLDGNRVKLDMTFAEGKPARVVDINI 189
DB 122 QYFNQATLNOAVAGLKEEYLGKGLNIQITPKVTKLARNRVDITIDEGSKAKITDIEF 181
QY 190 IGNOHFSADLDIVLAIKDKNI-NPLSKADRYTQEKLVTSLENLRAKYNLAGVFEIKD 248
DB 182 EGNQVYSRDKLRQMSLREGGIWTLTRSNQFNEQFAQDMKVTDFYQNNGYDFRILD 241
QY 249 AKLNINEDKNRIFRVEISLHGEQYRFGQTFGLNLT-YTQAELEALLKFAEGFSQAML 307
DB 242 TDIOQNEKTKOTIKITVHEGGRFWMGVKYSIEGDTNEVPKAELEKLLTMKPKWYERQOM 301
QY 308 EQTTNNISFKGDDGYVYQAQIRPVTRINDESRVDVEYIDPVHPVYVRRINFTGNFKTQ 367
DB 302 TAVLGEIQNRMGASGAYSEISVQPLPNAETKTVDVFLHIEPGRKIYVNEIHTGNKTR 361
QY 368 DEVLRRERMOLEGASLNQIOLSRALMRGTFKHVTVDRPVPNSPDQVDVNFVEEQ 427
DB 362 DEWVRRELQOMESAPYDTSKLQSKRVERVLLGYFDNVQFQDAVPLAGTDPKVDLNMSLTER 421
QY 428 PGSSTIAAGYSQSGVTFQFDVSONNFMGTGKHYNASFSESTREVSLSGMTNPFTVN 487
DB 422 STGSLDLGAGVQDFTGLVMSAGVSQDNLFCTGKSAALRASRSTKTLNGLSFTDPYFTAD 481
QY 488 GVSQSLSGYR--KTKYDNKNISNVLDYSYGSLSGYYPIDENQRIISFGLNADNTKLHGG 545
DB 482 GVSGLGYDYKAFDPRKASTSIKQYKTTTAGAGIRMSVPVTEYDRVNFGLVAEHLTVN-- 539
QY 546 RFMGISNVKQLMADGGKIQVDNNGIPDFKHDTYTNAILGNWYSSLDPRVPFTQGMHSV 605
DB 540 ---TYNKAPKHVADFIKKYKTDG--TDGSFGKWLKYGTVGWRNKRTDSALMPTRGYLTGV 595
QY 606 DLTGVF-GDK--THQKVYVQGGNIYRPIKKSVLK-----GYAKLGYG--NNLPFFYENFYAG 656
DB 596 NAEIALPGSKLQYYSATHNQWTFPPLSKTFTLMLGGEVGIAG-GYGRTKKEIPFFENFYGG 654
```



Query Match	25.0%	Score 1050;	DB 2;	Length. 792;
Best Local Similarity	31.5%;	Pred. No. 1.1e-49;		
Matches	258;	Conservative 159;	Mismatches 334;	Indels 68; Gaps 20;
Qy	23	STHAQAADFANDITITGLQRVTTSLSVLPRFGVVSENQLADGVKALYATGNFSDV	82	
	:	:	: : :           :	:

Db	14	TTTTAAFPFAKDIRVDGVGDLEQQIRASLPVRAQQRVTDNDVANIVRSFLVSGRFDDY	73
Qy	83	QVYHQEGRI--IYQVTERPLIAEINFEGNRRLIPKEGLEQGLKNAGLAVGQPLKQATVOMIE	141
Db	74	KA-HQEGDLVWSVWAKSIISDVKIKGNSIIPTEALKQNLDFANGFKVGDILIREKLNEFA	132
Qy	142	TELTNQIISOGYNTETTYKQTMLDGNRVKLTMTFAEGRPARVVDINIIGNHQHFSADLLI	201
Db	133	KSVKEHYASVGRYNATVEPTIVNTLPNNRAELIILQINEDDKAKIASLTFTKGNESVSSTLQ	192
Qy	202	DVLAIKDKNIPLSKADRYTOEKLVTSLENRAKRIYNAGVREIFEIKDAKLININEDKNRIF	261
Db	193	EQMELQDPDSMWKL--WGNMFCGAQFEKDLQAIRDYIYLNNGYAKAIIYKTVQVLNDEKTKVN	251
Qy	262	VEISLHEGEQYREFGOTQFLGNLTYYTQAELEALK-FKAEEGFSQAMLEQTTNNISTKFGD	320
Db	252	VTIDVNEGLQYDLRSARIICNLGMSAELEPILLSALHLNDFTRRSADIADVENAIAKALGE	311
Qy	321	DGYIIAQIRPVTRINDESRVDEVYIDIPVHVYVVRINPTGNFKTQDEVLRREMRQLEG	380
Db	312	RGYGTNTVNSVPDFDDANKTLAITFVVDAGRRUTVHLRFEQGTVSADSTLROEMRQEG	371
Qy	381	ALASNOKIQLSRARLMTGTFKKHVTVDTR--PVPNSPDQVDVNVFVEEOPSGSSTIAAGY	438
Db	372	TWNSQVLEUGKIRLDRTGTFE--TVENRIDPTNGSDEVDVYVYKERTWGTSINFGIGY	429
Qy	439	SQSGGVTFQFDVSONPMFTGKHVNASFSETREYVSLGMTNPYFTVNGVSQSLSGYYR	498
Db	430	GTESSISYQASVKQDNFLGTGAASVAGTAKNDYGTSVNLGYTEPYFTKDGV--SLGSGNVP	487
Qy	499	KTKYDNK---NISNVLDSTVGGSLSYGYPIDENQRIISFGL-----NADNTKHLHGGRPMGI	550
Db	488	FENYDNSKSTSSNYKRTTYGGSNVTLTGLFPVNNENNSYVVLGHTYNTKISNFALEYNRNLXI	547
Qy	551	SNVKQLMADGKKIQVDNNGIPDFEKHDYTTYNALILGNYSLSLDRPVPPTGCMHSVD--LT	608
Db	548	QSMK-----FKGNGIKTNDPDFS-----FGWYNISLNRGYPTPTGVKASLGGRV	592
Qy	609	VGFQDKTHQVYQGNITYRPF-----IKKSVLRYGAKLGYGN-NLPFVENFYAGGYGS	660
Db	593	IPGSDNKYKLSADVOGFIPLDRHLWVWSAKASGYAN-GFCGNKRLPFVQTYTAGIGS	651
Qy	661	VRGYDQSSLGPRQAYLTARRGQTTL-----GEVWGNALATFGSELILPLPKGDWI-D	715
Db	652	LRGFAYGISGPN-----IYQGONNFKNISSDVGIGNALATASAEIIVPTPFVSDKSON	706
Qy	716	QVRPVIEEGQVFDTT-GMDKQTDILTQPKDQATAEQNAKANRPLLTDQKQLRYASAG	774
Db	707	TVRTSLFVDAASVWNTKWKSDKNGLSNVLKD-----LPDYGKSSRIRASTG	753
Qy	775	VGATWYTPIGPLSISYAKPLNKKNQDQTDVQFOIGSVF	813
Db	754	VGFWQSPIGPLVFSYAKPIKKYENDDDVEQFQSIGGSF	792
RESULT	12		
Q9CJL1	ID	PRELIMINARY;	PRT; 791 AA.
AC	Q9CJL1		
DT	01-JUN-2001	(TEMBLrel. 17, Created)	
DT	01-JUN-2001	(TEMBLrel. 17, Last sequence update)	
DE	01-DEC-2001	(TEMBLrel. 19, Last annotation update)	
DE		HYPOTHETICAL PROTEIN PM1992.	
GN	PM1992		
OS		Pasteurella multocida.	
OC		Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;	
OC		Pasteurella.	
OC		NCBI_TaxID=747;	
RN		[1]	
RP		SEQUENCE FROM N.A.	
RC	STRAIN=PM70;		
RX	MEDLINE=21145866; PubMed=11248100;		
RA	May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;		



```
Db 477 SLGGNVSEEDSKSNTSAGYRYSYGGNLTGLFPVNNENSYLGVGYTNKL----- 530
QY 549 GISNV-KOLMADGGKIQVNDNNGIPDFK-HDYTTYNAILGNWYSSSLDRPVPFTQGMHSHVD 606
Db 531 --KNTAPEVNRDLYRQSMKYNDSWTFKSHD---FDLSFGWNYSNLNRGYFPTKGYRANIG 585
QY 607 --LTVGFGDKTHOKVYVYQGYNIYRPF-----IKSVLRGVAKLYGNNLPFYENF 653
Db 586 GRVTIPGSDNKKYKLNAAEQGYPLDREHGWVLSRSISASFADGFS---GKRLPFYQY 641
QY 654 YAGGVSVRGVDOSSLSGRPSQAYLTARRGQQTLL--GEVVGGNALATGSELILPLPKG 711
Db 642 SAGGSGLSRGFRAYGAPGNA-IYTRQCPDPSVCLVSSDVIGGNAMVTASTELIVPTPEVA 700
QY 712 DW-IDQVRPVIFIEGGQVFDTTGMDKQITDLTQFKDPQATAPQNAKAA--NRPLLTQDQK 768
Db 701 DKNQNSVRTSLFVDAASVWNT-----RWKAEDKAKFAKLNVPDYSDPSR 744
QY 769 LRSAGVGTATWTPIGPLSISYAKPLNKONDQDTDTVQFQIGSVF 813
Db 745 VRASAGVALOWQSPICPLVFSYAKPLKRYQGGDEIQFQFSIGGTF 789

RESULT 14
Q9ZE03 PRELIMINARY; PRT; 768 AA.
AC Q9ZE03;
DT 01-MAY-1999 (TremBLrel. 10, Created)
DT 01-MAY-1999 (TremBLrel. 10, Last sequence update)
DE 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE OUTER MEMBRANE PROTEIN OMPI (OMPI).
GN RPI160.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsieae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=MADRID E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria.";
RL Nature 396:133-140(1998).
DR EMBL: AJ235270; CAAL4627.1; -
DR InterPro: IPR000184; Bac_surfaG_d15.
DR Pfam: PF01103; Bac_surfaceAg; 1.
KW Complete proteome.
SQ SEQUENCE 768 AA; 87196 MW; BIAB12D783D9FDCB CRC64;

Query Match 17.98; Score 754; DB 16; Length 768;
Best Local Similarity 26.38; Pred. No. 2.le-33;
Matches 216; Conservative 145; Mismatches 383; Indels 76; Gaps 20;

QY 7 KGFQVSAMTAVMMVMSTHAQAADPMANDITITGLQRTVIESLQVLPFRLGQVYSENL 66
Db 2 KIISIKTLILLIFHYHISFADYVIRKIYIEGHNHVRSTIESYKLNVGTEYNNKSKE 61
QY 67 ADGVKALYATNFSDVQVY-HOEGRIYQVTERPLIAEFNFGNRLIPKEGLQGLKNAG 125
Db 62 DEAIKRLYATSLFRNINMYITDGNLIVNVTETPTPSSVFSNGSKIRTNILAKEIYMS 121
QY 126 LAVQPLQAQVQVMEIETLTNQYISQGYNTIEITVQKQMTLQGNRVKLDMTAEGKPARV 185
Db 122 ---GESLSQALIEDVKKILEIYRSRGFSTKVTPKIKSLNNRVKVFIDFAEGPKTVIK 178
QY 186 DINTIGNHFSADLIDVLATKDKNK-INPLSKADRYTOEKLVTSLNLRAYLNAGFYRF 244
Db 179 SIYFSGNEHYSDELKSLVLTKEKRWFRFLESNDTDPDRVEYDKELLREFYQSVGFADF 238
```

```
QY 245 EIKDAKLINEDKNRIFVEISLHGEQYRFGQTFGLGNLTYYQ-AEALALKKAEGBGFS 303
Db 239 RVISASVALNDTKYETITYSIEGEKRYFGNVTIDNKLINIKLNKIVNIKQGRIFN 298
QY 304 QAMLEOTTNNISTKFGDDGYYIAIRPVTRINDESRVTDVEYIDVPVYVVRINPTGN 363
Db 299 MKTVDDIAEKIGBYFTANGYPAYNVYPDIKKND-NHTADIKFTIEKADKYIYINKININ 357
QY 364 FKTQDVLRREROLGALASNOIKIOLSRARLMRTGFFKHVTVTDTRPVNPSPDQVDVNEF 423
Db 358 LKTEDHIVIRAKIEGDMVNRYSIEKGERNLRLDYEKVSISLAQT-KAKDKYDVNVE 416
QY 424 VEEQPSGSTIIAAGYSQSGVTFQFDVSONFMGTGKHVNASFSRSRETVYSLGMTNYP 483
Db 417 VDEKSTSSIGFDLGYNTAGLFGFRSFLERNLVGTCKLLNAGVQVSKNSTSYGCTDPH 476
QY 484 FTVNGVSQSLSGYY----RKYDKNKINISYVLDSDGGSLSYGYPTDE--NQRISFGLNA 537
Db 477 FLDRDLSSVNAFRNVTGRGASVINTDQSYKLHSIGVKISLGYDMKEDLSHEIDYLKR 536
QY 538 DMTKLGHRFMGISNVKQLMADGGKIQVNDNNGIPDFKHDVTTYNAILGNWYSSSLDRPVP 597
Db 537 DILSAPSP-----SNSIFLNEQMKLITSAG-----HTIT-----YDQDNKIVP 577
QY 598 TQG--MSHVDLTGVFGDKTHQKVYVQGYNIYRPFIRKSVLRGYAKLG-----YGNLPPF 649
Db 578 KNGYLVSQTQEFAGVGGDNKYIKHEIECKFYKSFINNKYTKLSAAGDMAGLGGKMWRI 637
QY 650 YENFVAGGVSVRGVDOSSLSGRPSQAYLTARRGQQTTLGEVVGGNALATGSELILPLPF 709
Db 638 SDRFNLDGY-SLRGFASGGVGRP-----EKNTEGLEGGERYYTFTSTELNFTPV 685
QY 710 KGDVTDQVRPVIFIEGGQVFDTTGMDKQITDLTQFKDPQATAPQNAKARNPLLTQDKOL 769
Db 686 PEEF-NFTGAVFIDLGSVAG-VGLNKK-----QYTPNG-----FYNDQSL 724
QY 770 RYSAGVGATWYTPIGPLSISYAKPLNKKNQNDQDTVQFQI 809
Db 725 RASVGFGEFTWTRFAPIRMDWGFPIKKQYD--DTQNFHL 762

RESULT 15
Q92Q48 PRELIMINARY; PRT; 776 AA.
AC Q92Q48;
DT 01-DEC-2001 (TremBLrel. 19, Created)
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)
DE PUTATIVE OUTER MEMBRANE TRANSMEMBRANE PROTEIN.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21368234; PubMed=11474104;
RA Galibert F., Finan T.M., Long S.R., Puehler A., Abola P., Ampe F.,
RA Barloy-Hubler F., Barnett M.J., Becker A., Boistard P., Bothe G.,
RA Boutry M., Bowser L., Buhrmester J., Cadieu E., Capela D., Chain P.,
RA Cowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F.,
RA Gloux S., Godrie T., Goffeau A., Golding B., Gouzy J., Gurjal M.,
RA Hernandez-Lucas I., Hong A., Huizar L., Hyman R.W., Jones T., Kahn D.,
RA Kahn M.L., Kalman S., Keating D.H., Kiss E., Komp C., Lelaure V.,
RA Masuy D., Palm C., Peck M.C., Pohl T.M., Portetelle D., Purnelle B.,
RA Rasperger U., Surzycki R., Thebault P., Vandenbol M.,
RA Vorhoeiter F.J., Weidner S., Wells D.H., Wong K., Yeh K.-C., Batut J.;
RT "The composite genome of the legume symbiont Sinorhizobium meliloti.";
RL Science 293:668-672(2001).
DR EMBL: AL591787; CAC46081.1; -.
KW Complete proteome.
SQ SEQUENCE 776 AA; 84520 MW; CE261E6372A94726 CRC64;
```

Query Match 17.8%; Score 748.5; DB 16; Length 776;  
Best Local Similarity 25.5%; Pred. No. 4.2e-33;  
Matches 216; Conservative 152; Mismatches 354; Indels 125; Caps 20;

```
QY 11 VSAMTAMVMVM-----STHAQAADFMANDITITGLQRYTIESLSQSLVLPFRLGQV 60
Db 11 VSAFALSASWATGTGVLVAGTSAQA--VINRVEVRGATRVSAETVRANITIVPKS 68

QY 61 VSENQLADGVKALYATNFSQVYHQEGRIIYQVTERPLIAEINFEGNRLIPKEGLQEG 120
Db 69 FSNADIDASVKRLYATGYSITVSGSLVSVSENQLVNVQVFNNGRKKDKKLGQV 128

QY 121 LKNAGLAVGQPLKQATVQMIETELTNQYISQYYNTEITVKQTMLDGNRVKLDMTFAEGK 180
Db 129 VRTQPLG---PYSEATVETDQAIQIRDAYAAIGRSDDVTVTQVVPVIAEGRVNLAFVINEGE 185

QY 181 PARVVDINIIGNQHFSDADLDVLAIKDNKI-NPLSKADRYTQEKLVTSLENLRAKYLNA 239
Db 186 RTKITQINFVGNVEISGRVLSQSVIATKESGIFSELTTRDVTNPDKLRADPELLRQFYNR 245

QY 240 GFVRFEIKDAKLINEDKNRIFVEISLHEGEQYRFQGTQFLGNLTYYQA-ELEALLKFKA 298
Db 246 GYADFQVVSSEALNEATNEVTVTITIEGPRYDFGPNVIESTVEGIDAEELRGLVQSRE 305

QY 299 EEGFSQAMLEQTTNNISTKFGDDGYIAQIRPVTRINDESKTVDVVEYYIDPVHPVYVRR 358
Db 306 GTVYKAKDIQSTMSEISKRVASEGYPARVTPRGNRDLNHTIAVDYLVQDGERAYVERI 365

QY 359 NFTGNFTQDEVLRRERQLEGALASNQIKLSRARLMRTGFEKHKVTDTRPVNSPDQV 418
Db 366 EIRGNTTRDYVIRREFDVGEQDFAFQEMVARAKRRLREALGYFSSVNIQTQ-GSAADRV 424

QY 419 DVNFVVEEQSPGSGSTIAAGYS--QSGGVTEQFDVSNFMGTGKHVNASFSSRSETREYVS 476
Db 425 VIVVDVQDQSTGFGIGAGYSAGDGGGFLVEASTEEKNFLGRQYIRLAAGKGEDSQTYN 484

QY 477 LGMTNPFYTVNGVSQSLSGYYRKYTKYDNKNISNVVLDVSYGGSLSYGYPIDENORISFGLN 536
Db 485 VSFTPEYFLGYRLAAGFDLEKFNENDEDD--DNYSYNDQGFSLRVATPITENLSTTLRYN 541

QY 537 ADNTKLHGGRFMGISNVKQLMADGKIQVDNNGIPDFKHDTYTYNAILGNWYSSLDPRVF 596
Db 542 YTELEFGDR-DELSPPYDRVIGS-----PWRSSISQSITYNFLDDAQL 586

QY 597 PTQGMHSVDLT-VGFGDKTHQKVYQGNTRYRPFIRKSVLRGYAKLG-----643
Db 587 PHEGILASVTQEFAGLGTS-----DFYK-----LTGKAKWYITLHDEADIIIGSL 631

QY 644 -----GNLPPFENFYAGGYSVGRYDQSSSLGPRSQAYLTARRGOOTLGEVVGG 693
Db 632 SGSAGHLFETSGSLEFVDFQFLNS-NDIRGFERNGLGPRMNN-----GDALGG 678

QY 694 NALATFGSELILPLP-----FKGDWIDQVRPVIFIEGGQVFDTCGMKQTDIDLTQFKD 746
Db 679 TTYFTASAEATFPLGLPRDSGFRG-----ALFVDACTLYG-----ND 716

QY 747 PQATAEQNAKAAANRPLLTQDKQLRYSAQVATWYTPIGPLSISYAKPLNKKQNDQTDTVQ 806
Db 717 VEIGPGESVRGDN-----ASLRASVGVSLIWASPEGPLRVDYAVFVAKEDFDEVQNFK 769

QY 807 FQIGSVF 813
Db 770 FGINSSF 776
```

Search completed: September 5, 2002, 10:03:04  
Job time: 538 sec

PD 5/3/99

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 09:56:07 ; Search time 61.46 Seconds  
(without alignments)  
1469.297 Million cell updates/sec

Title: US-09-701-711-2  
Perfect score: 813  
Sequence: 1 MRSYFKGFQVSAMTMMVM.....LNKKQNDQTDVQFQIGSVF 813

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Word size : 15

Total number of hits satisfying chosen parameters: 4

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : A\_Geneseq\_032802.\*

- 1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	813	100.0	813	21	AA1980
2	749	92.1	813	21	AA1981
3	16	2.0	17	21	AA1983
4	15	1.8	16	21	AA1984

ALIGNMENTS

RESULT 1  
AA1980

ID	AA1980	standard; Protein; 813 AA.
XX	AA1981	
AC	AA1982	
XX	AA1983	
DF	14-MAR-2000	(first entry)
XX	M. catarrhalis BAS027	polypeptide #1.
DE	BAS027	OMP85; outer membrane protein; otitis media; treatment;
XX	BAS027	diagnosis; bacterial infection.
KW	Moraxella catarrhalis.	
XX	WO9963093-A2.	
PN	WO9963093-A2.	
XX	09-DEC-1999.	
PD	09-DEC-1999.	
XX	31-MAY-1999;	99WO-EP03822.
PF	31-MAY-1999;	99WO-EP03822.
XX	03-JUN-1998;	98GB-0011945.
PR	03-JUN-1998;	98GB-0011945.
XX	08-MAR-1999;	99GB-0005304.
PR	08-MAR-1999;	99GB-0005304.
XX	(SMIK ) SMITHKLINE BEECHAM	BIOLOGICALS.
PA	(SMIK ) SMITHKLINE BEECHAM	BIOLOGICALS.
XX	Vinals-Bassols C;	
PI	Vinals-Bassols C;	
XX	WPI; 2000-105700/09.	
DR	N-PSDB; AA229550.	
DR	N-PSDB; AA229550.	
XX	Novel BAS027 polynucleotide and polypeptides from Moraxella	catarrhalis useful for treating M. catarrhalis infection such as otitis
PT	media	
PT	media	
XX	Claim 1; Fig 3; 109pp; English.	
PS	Claim 1; Fig 3; 109pp; English.	
XX	The present sequence is BAS027 polypeptide, which shows significant	homology to Neisseria meningitidis OMP85 outer membrane protein. It is
CC	homology to Neisseria meningitidis OMP85 outer membrane protein. It is	encoded by DNA obtained from chromosomal DNA library of Moraxella
CC	catarrhalis strain Mc2931 (ATCC 43617). BAS027 polynucleotide and	polypeptide can be used for diagnosis and staging of disease, determining
CC	polypeptide can be used for diagnosis and staging of disease, determining	susceptibility to a disease and to prepare medicaments for treating M.
CC	susceptibility to a disease and to prepare medicaments for treating M.	catarrhalis infections, especially otitis media. The BAS027 DNA can be
CC	catarrhalis infections, especially otitis media. The BAS027 DNA can be	used as probe for screening of genetic mutations, serotype, taxonomic
CC	used as probe for screening of genetic mutations, serotype, taxonomic	classification or identification. BAS027 agonists, antagonists and
CC	classification or identification. BAS027 agonists, antagonists and	antibodies may be used to prevent and/or treat bacterial infections.
XX	antibodies may be used to prevent and/or treat bacterial infections.	
SQ	Sequence	813 AA;
Query Match	100.0%; Score 813; DB 21; Length 813;	
Best Local Similarity	100.0%; Pred. No. 0;	
Matches	813; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	1 MRSYFKGFQVSAMTMMVMSTHAQAADFMANDDITGLQRTVIESLQSVLPFRLQGV	60
Db	1 mrsyfkfgqvsamtmmvmsthaqaadfmnditglqrvtieslqsvlpfrlqgv	60
Qy	61 VSENLADGVKALYATGNGFSDVQVYHQBGRYIYQVTERPLIAEINFEGRNLPKEGLQEG	120
Db	61 vsenladgvkalyatgngfsvdqvyhqbgrlyqvtterpliaeinfegnrlpkeglqeg	120
Qy	121 LKNAGLAVGQPLKQATVOMIETLNTQYISQGYNYNTEITVKTMLDGNRKLDMTFAEGK	180
Db	121 lknaglavgqplkqatvqmieteltnqyisqgynteitvktmldgnrvklmdtfaegk	180
Qy	181 PARVVDINIINQHFSDADLDVLAIKDNKINPLSKADRYTQEKLVTSLENIRAKYLNAG	240
Db	181 parvvdiniinqhfdsadldvlaikdnkinplskadrytqeklvtslenirakyl nag	240
Qy	241 FVRFEIKAKLNINEDKNRIFVEISLHEGEYRFQGTQFLGNLTYYTQAEALLKFAEE	300
Db	241 fvrfeikaklninedknrifveislhegeyrfqgtqflgnltyyqaeallkfaee	300

QY 301 GFSQAMLEQTTNNISTKFGDDGYIAQIRPVTRINDESRTVDVEYIDPVHPVYVRRINF 360  
 Db 301 gfsqamleqttnnistkfgddgyiaqirpvtrindesrtvdveyidpvhpyvrrinf 360  
 QY 361 TGNFKTQDEVLRERMQLEGALASNQIQLSRARLMTGFFKHVTVDTTRVPNPSDPOVDV 420  
 Db 361 tgnfktdqdevlrermqlegalasnqqlsrrarlmtgffkhvtdtrvpnpdpodv 420  
 QY 421 NFVVEEQPSGSSTIAAGYSQSGGVTQFDVSONNFMTGKHVNASFSRSTRETVYSILGMT 480  
 Db 421 nfveeqpsgsstiaagysqsggvtfqdvsnnfmgtkhvnasfsrsetrevysilgmt 480  
 QY 481 NPYFTVNGVSQSLSGYYRKTGYDNKNISNVLDYSGSLSGYPIDENQRIISFGLNADNT 540  
 Db 481 npyftvngvsqslsgyyrktgynknisnvyldysgslsygypidenqrisfglnadnt 540  
 QY 541 KLHGGRFMGINSVKQLMADGGKIQVDNNGIPDFKHDTTYNAILGNWYSSLDPRVPPTQG 600  
 Db 541 klhggrfmginvskqlmadggkigvdnngipdfkhdtytnailgnwnyssldprvpptqg 600  
 QY 601 MSHSVDLTVGFGDKTHOKVYQGNIRPFTKKSVLRGYAKLGYGNLUPFENFYAGSYGS 660  
 Db 601 mshsvdltvfgdkthokvyqgnirpftkksvlgryaklgygnlupfenfyagsgys 660  
 QY 661 VRGYDQSSLGPRSQAYLTARRGQOTTLGEGVVGGNALATFGSELILPLPFKGDWIDQVRPV 720  
 Db 661 vrgydqsslgprsqayltarrgqottlgegvvggnalatfgselilplpfkgdwidqvrpv 720  
 QY 721 IFTEGGVFTTGMDKQTTIDLTQFKDQATAEONAKAANRPLLTQDKOLRYSAGVGATWY 780  
 Db 721 ifteggvfttgmdkqttidltqfkdpqataeqnakaaanrpiltdqkqlrysgvgatwy 780  
 QY 781 TPIGPLSISYAKPLNKKNDQTTDTPVQFQIGSVF 813  
 Db 781 tpigplsisyakplnkkndqtdtpvfqigsvf 813  
 RESULT 2  
 AAY44391  
 ID AAY44391 standard; Protein; 813 AA.  
 XX AC AAY44391;  
 XX DT 14-MAR-2000 (first entry)  
 XX DE M. catarrhalis (ATCC 43617) BAS027 polypeptide #2.  
 XX KW BAS027; OMP85; outer membrane protein; otitis media; treatment;  
 XX KW diagnosis; bacterial infection.  
 XX OS Moraxella catarrhalis.  
 XX PN W09963093-A2.  
 XX PD 09-DEC-1999.  
 XX PF 31-MAY-1999; 99WO-EP03822.  
 XX PR 03-JUN-1998; 98GB-0011945.  
 XX PR 08-MAR-1999; 99GB-0005304.  
 XX PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 XX PI Vinals-Bassols C;  
 XX DR WPI: 2000-105700/09.  
 XX DR N-PSDB; AAZ29551.  
 XX PT Novel BAS027 polynucleotide and polypeptides from Moraxella  
 XX PT catarrhalis useful for treating M. catarrhalis infection such as otitis  
 XX PT media  
 XX Claim 1; Page 102-104; 109pp; English.  
 PS

XX The present sequence is BAS027 polypeptide, which shows significant  
 CC homology to Neisseria meningitidis OMP85 outer membrane protein. It is  
 CC encoded by DNA obtained from chromosomal DNA library of Moraxella  
 CC catarrhalis strain Mc2931 (ATCC 43617). BAS027 polynucleotide and  
 CC polypeptide can be used for diagnosis and staging of disease, determining  
 CC susceptibility to a disease and to prepare medicaments for treating M.  
 CC catarrhalis infections, especially otitis media. The BAS027 DNA can be  
 CC used as probe for screening of genetic mutations, serotype, taxonomic  
 CC classification or identification. BAS027 agonists, antagonists and  
 CC antibodies may be used to prevent and/or treat bacterial infections.  
 XX Sequence 813 AA;  
 SQ  
 Query Match 92.1%; Score 749; DB 21; Length 813;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 749; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 65 QLADGVKALYATGNFSDVQVYHQBGRILYQVTERPLIAEINFEGRNRLIPKEGLOEGLKNA 124  
 Db 65 qladgvkalyatgnfsdvqvyhqbgrilyyqvterpliaeinfegnrlipkeglqeglkna 124  
 QY 125 GLAVGOLPKQATQOMIETELTNQYISQGYNTETTVKQTMLDGNRVKLDMTFAEGKPARV 184  
 Db 125 glavgolkqatqomieteltnqyisqgyntetvtvkqtmldgnrvklmdmtfaegkparv 184  
 QY 185 VDINIIGNHFSDADLIDVLAIKDNKINPLSKADRYQEKLVTSLENLRAKYLNAGVRF 244  
 Db 185 vdiniignhfsdaddlidvlaikdnkinplskadryqeklvtslenlrakylnagvrf 244  
 QY 245 EIKDAKUNINEDKNRIFVEISLHEGEQYRFGOTQFLGNLYTQAEALALLKFAEEGFSQ 304  
 Db 245 eikdakuninedknrifeislhegeqyrfgotqflgnlytqaealallkfaeeegfsq 304  
 QY 305 AMLEQTTNNISTKFGDDGYIAQIRPVTRINDESRTVDVEYIDPVHPVYVRRINFNGNF 364  
 Db 305 amleqttnnistkfgddgyiaqirpvtrindesrtvdveyidpvhpyvrrinfngnf 364  
 QY 365 KTQDEVLRERMQLEGALASNQIQLSRARLMTGFFKHVTVDTTRVPNPSDQVDVNFV 424  
 Db 365 ktqdevlrermqlegalasnqqlsrrarlmtgffkhvtdtrvpnpdpdvnfv 424  
 QY 425 EEQPSGSSTIAAGYSQSGGVTQFDVSONNFMTGKHVNASFSRSTRETVYSILGMTNPFY 484  
 Db 425 eeqpsgsstiaagysqsggvtfqdvsnnfmgtkhvnasfsrsetrevysilgmtnpfy 484  
 QY 485 TVNGVSQSLSGYYRKTGYDNKNISNVLDYSGSLSGYPIDENQRIISFGLNADNTKLHG 544  
 Db 485 tvngvsqslsgyyrktgynknisnvyldysgslsygypidenqrisfglnadntklhg 544  
 QY 545 GREFGINSVKOLMADGGKIQVDNNGIPDFKHDTTYNAILGNWYSSLDPRVPPTQGMKSHS 604  
 Db 545 grfmgisnvkqlmadggkigvdnngipdfkhdtytnailgnwnyssldprvpftqgmshs 604  
 QY 605 VDLTVGFGDKTHOKVYQGNIRPFTKKSVLRGYAKLGYGNLUPFENFYAGSYGVRGY 664  
 Db 605 vdltvfgdkthokvyqgnirpftkksvlgryaklgygnlupfenfyagsgyvrsgy 664  
 QY 665 DQSSILGPRSQAYLTARRGQOTTLGEGVVGGNALATFGSELILPLPFKGDWIDQVRVIFIE 724  
 Db 665 dqssilgprsqayltarrgqottlgegvvggnalatfgselilplpfkgdwidqvrpvif 724  
 QY 725 GGQVFDFTGMDKQTTIDLTQFKDQATAEONAKAANRPLLTQDKOLRYSAGVGATWYTFIG 784  
 Db 725 ggqvfdftgmdkqttidltqfkdpqataeqnakaaanrpiltdqkqlrysgvgatwytfig 784  
 QY 785 PLSISYAKPLNKKNDQTTDTPVQFQIGSVF 813  
 Db 785 plsisyakplnkkndqtdtpvfqigsvf 813  
 RESULT 3



AA44393  
ID AAY44393 standard; peptide; 17 AA.  
XX  
AC AAY44393;  
DT 14-MAR-2000 (first entry)  
XX  
DE Moraxella catarrhalis BASB027-specific synthetic peptide #2.  
XX  
KW BASB027; OMP85; outer membrane protein; otitis media;  
KW bacterial infection.  
XX  
OS Synthetic.  
XX  
PN WO9963093-A2.  
XX  
PD 09-DEC-1999.  
XX  
PF 31-MAY-1999; 99WO-EP03822.  
XX  
PR 03-JUN-1998; 98GB-0011945.  
PR 08-MAR-1999; 99GB-0005304.  
XX  
PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
XX  
PI Vinals-Bassols C;  
XX  
DR WPI; 2000-105700/09.  
XX  
PT Novel BASB027 polynucleotide and polypeptides from Moraxella  
PT catarrhalis useful for treating M. catarrhalis infection such as otitis  
PT media -  
XX  
PS Example 8; Page 64; 109pp; English.  
XX  
CC The present sequence is a Moraxella catarrhalis BASB027  
CC polypeptide-specific peptide which can be coupled to KLH and used for  
CC producing antibodies against BASB027. Anti-BASB027 antibodies can be used  
CC to treat bacterial infections.  
XX  
SQ Sequence 17 AA;

Query Match 2.0%; Score 16; DB 21; Length 17;  
Best Local Similarity 100.0%; Pred. No. 1e-08;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 676 YLTARRGQQTTLGEVV 691  
Db 1 yltarrgqqtllgevv 16  
|||||

RESULT 4  
AAY44392  
ID AAY44392 standard; peptide; 16 AA.  
XX  
AC AAY44392;  
DT 14-MAR-2000 (first entry)  
XX  
DE Moraxella catarrhalis BASB027-specific synthetic peptide #1.  
XX  
KW BASB027; OMP85; outer membrane protein; otitis media;  
KW bacterial infection.  
XX  
OS Synthetic.  
XX  
PN WO9963093-A2.  
XX  
PD 09-DEC-1999.  
XX  
PF 31-MAY-1999; 99WO-EP03822.  
XX

PR 03-JUN-1998; 98GB-0011945.  
PR 08-MAR-1999; 99GB-0005304.  
XX  
PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
XX  
PI Vinals-Bassols C;  
XX  
DR WPI; 2000-105700/09.  
XX  
PT Novel BASB027 polynucleotide and polypeptides from Moraxella  
PT catarrhalis useful for treating M. catarrhalis infection such as otitis  
PT media -  
XX  
PS Example 8; Page 64; 109pp; English.  
XX  
CC The present sequence is a Moraxella catarrhalis BASB027  
CC polypeptide-specific peptide which can be coupled to KLH and used for  
CC producing antibodies against BASB027. Anti-BASB027 antibodies can be used  
CC to treat bacterial infections.  
XX  
SQ Sequence 16 AA;

Query Match 1.8%; Score 15; DB 21; Length 16;  
Best Local Similarity 100.0%; Pred. No. 1.1e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 790 YAKPLNKKQNDQTD 804  
Db 2 yakplnkqndqtdt 16  
|||||

Search completed: September 5, 2002, 10:04:37  
Job time: 510 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 09:59:52 ; Search time 25.36 seconds  
(without alignments)  
783.044 Million cell updates/sec

Title: US-09-701-711-2  
Perfect score: 813  
Sequence: 1 MRNSYFKGFQVSAMTAVMM.....LNKKQNDQDTVQFQIGSVF 813

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 231628 seqs, 24425594 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
-----						

No matches found

Search completed: September 5, 2002, 10:05:08  
Job time: 316 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 10:03:07 : Search time 39.77 Seconds  
(without alignments)  
1964.310 Million cell updates/sec

Title: US-09-701-711-2  
Perfect score: 813  
Sequence: 1 MRNSYFKGFQVSAMTAVMM.....LNKKQNDQDTVQFQIGSVF 813

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : PIR\_71.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
-----						

No matches found

Search completed: September 5, 2002, 10:05:54  
Job time: 167 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 10:05:57 ; Search time 20.98 Seconds  
(without alignments)  
1500.429 Million cell updates/sec

Title: US-09-701-711-2  
Perfect score: 813  
Sequence: 1 MNSYFKGFQVSAMTMAVM.....LNKKQNDQDTVQFQIGSYF 813

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
-----					

No matches found

Search completed: September 5, 2002, 10:12:48  
Job time: 411 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 10:05:12 ; Search time 63.51 Seconds  
(without alignments)  
2214.531 Million cell updates/sec

Title: US-09-701-711-2  
Perfect score: 813  
Sequence: 1 MRNSYFKGQVSAMTAVMM.....LNKKQNDQTDVQFIGSVF 813

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SPTREMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp Vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

No matches found

Search completed: September 5, 2002, 10:12:20  
Job time: 428 sec



QY	68	DGVKALYATGNFSDVQVYHOEGRILIIQTVERPLIAEINTEGKNRLIPKEGLQGSLKNAGLA	127
	:	: :     :        :	:
Db	61	NTRALFATGNFDVRVLRGNTLLVQVKERPTIASITFSGNKSXKDDMLKNLEASGYR	120
QY	128	VGOPLKQATQMIEITLNOYISQVYNTEITWKQPMLDGNVKLDMTFAEGKPARVVDI	187
Db	121	VGESLDRTTILSDIEKGLEDFFYSVGKYSASVAVVTPLPNRVDLKLVOEGVSAKIQOI	180
QY	188	NIIGNQHFSADLIDLVAIKD-----NKINPLSKADR-YTOEKLTYSLENRAKYLNAGF	241
Db	181	NIIVNHAFTSEELISHQLRFDEVPWNVV-----GDQRKOKLAGDELTLTASYLYDRGY	235
QY	242	VREIFEIDAKLININEKDNRIFVEISLHEGBOYRFGQOFLGNLTYYTOAELEALLKPKAEBG	301
Db	236	ARENIDSTQVSLTPDKKGIYIWTNEGDOYKLSGVQVSGNLASHSAETKUTKEPGBL	295
QY	302	FSOAMLEQTTNISTKFGDDYYYAQIRPVTRIINDESRTVDVEYYITDPVHPHYVRRINF	361
Db	296	YNGTKYTKMEDDTFKLLGRYGAYPRVQSOPPEINDADKTVKLRNVNDACNREFYKRIRFE	355
QY	362	GNEPTODEVLIREMROLEGALASNOKILOLSRARLMRTGFKHVTVTDTRVPVNSPOQDVN	421
Db	356	GNDTSKSVLRRMRQMEGANWGLSDLVDOGKERLNLFGEFVEDTDOVRPGSPQOVVV	415
QY	422	FVVEEOPSGSSTIAAGYSOSGGVTFQFYSONFMGTGKHVNASFSSRETREVYSLGMTN	481
Db	416	YKVKERNTGSFNFGICYGTGESGVFOAGVQQDNWLGTGYSVINGTKNDYQIYSELVTN	475
QY	482	PFTVNGVSQSLSYYRKTYDNKNISNVLSYSGSLSYGYPIDENORISFGLNADNFK	541
Db	476	PYFTVDVSLGORIFYNPDEFADADLSDTNKSXYGTDVTLGPPINEYNTLRAGLYVHNK	535
QY	542	LHGREGMGISNWK-QUMADGKITVDNNIGIPDFKHDTYTNAILGNWYSLSLDRVPFPQOG	600
Db	536	L-----SNMQPOLAMDRYLESMDPDPAADFETFN--YGWTYNKLDRGYPTDG	585
QY	601	--MSHSDVLTUVGGRTHOKVYVQGINYPFTK--KSVLRGYAKILGYGNL-----PYFE	651
Db	586	SRYNLTKGVTIPIGSDNEYTKVSLDTATYVIPIDNDHKWVYLGRTRMGYGDLGGKEMPEFY	645
QY	652	NFYAGYGVSRVYDQSSLGRSQAYLTARRGOOTT-----LGVEVGN	694
Db	646	NFYAGSSTVRGFSQSWTIGPKA----VYNKAHTSHWDDDDYEDCTQESGCKSDDAVGN	701
QY	695	ALATFGEILPLPFKGD-WIDQVRPIVIEGGQVFDTT-GMDKOTIDLTQFKDPOATAE	752
Db	702	AMAVASLEFITPTFISEKYANSVRTSFWDGMGTVMYDINWDPSSAPSDVPDYSDP-----	756
QY	753	QNAKAANRPLLTDQKOLRYSAGVATWYTPIGLSTISYAKPLNKKONQOTDITVQOIGSV	812
Db	757	-----GNIRMSAGIALOWMPSPLGVFYSYAQPFPKYDGDKAEOFQFNIGHT	802
QY	813	F	813
Db	803	W	803

```

RESULT      4
A64742     hypothetical protein b0177 - Escherichia coli
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 29-Sep-1997
C:Accession: A64742
A:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;
A.; Rose, D.J.; Mau, B.; Shao, Y.
C:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: A64742
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-810 <BLAT>

```

A: Cross-references: GB:AE000127; GB:U00086; NID:g1786370; PIDN:AAC73288.1; PID:g1786370  
A: Experimental source: strain K-12, substrain MG1655  
C: Superfamily: protective surface antigen D-15

Query Match	28.8%;	Score 1210;	DB 2;	Length 810;
Best Local Similarity	34.0%;	Pred. No. 3.9e-62;		
Matches 287;	Conservative 149;	Mismatches 330;	Indels 78;	Gaps
QY 14	MTAVMMVMS-----THAAADFAMANDITITGLQRTVITSLQSVLPERLGOVSVENQLA 67			
DB 1	MAKKLLIASLLSFSSATVYGAEGFVKDTHFEGLRQVAVGALLSPVTRGDTVDNDEDIS 60			
QY 68	DGVKALYATGNFSDVQVYHOEGRIIYQVTERPLIABINFEGNRLIPKEGLOGLKNAGLA 127			
DB 61	NTIRALFATGNFEDVRVLRGDFTLLVQVKERPRTIASITTSGNKSVKDDMLKQNLASGYR 120			
QY 128	VGOPLQKATVOMLETETLNOYISOGYYNTEITVQKPMLOGNRVYKLDMTAEGKPARVDI 187			
DB 121	VGESLDRTITADIEKLEDDFYISVGKYSASVKAIVTPLPRNRVDLKLVOEGVSABIQI 180			
QY 188	NIIGNQHFSADALIDVLAIKD-----NKNPLSKADR-VTQEKLVTSLEMLRAKYLNAGF 241			
DB 181	NIVGNHAFVTDLISHFQLRDVEVPWNV-----GDRKYQOKLAGDLETLSYYLDRGY 235			
QY 242	VREIKDAKLINEDKNRIFVEISLHEGQYRFGQFQFLGNLTYYTOAELEALLKFKAEES 301			
DB 236	ARNIDSTQVSLRPDKKGIYVTVNITGDOYKLSGVGEVSGNLASHSAETQUTKIEPGL 295			
QY 302	FSQAMLEOTNNITSTKFGDDGYAAGIAPVPRINDESRTVDVEYITDPVHPVYVRINFT 361			
DB 296	YNGTKVTKMEDDILKLLGRYGIAYPRVQSPPEINDADKTVKLVNVDAGNRFVVKIRFE 355			
QY 362	GNPKTQDEVLRREMOLEGALASNQKIQLSARLMTGFFKHVYVDTRPVPNSPDQVDYN 421			
DB 356	GNDSKDAVLRREMOEGALSDLVDOCKERLNLGFFETVDTQVRPGSPDQVDV 415			
QY 422	FVVEEOPSGSSTIAAGYSQSQGYTFDFDYSONNFMCTGKHVNASFSRSSTREVISLGMTN 481			
DB 416	YKVKERTNGFNGIGYGESGVSPQAGVOODNWLGTGAVGINTGKNDIOTYAEULSVTN 475			
QY 482	PYFTVNGVSOSLGYRKTKYDNKNKINSYVLDSYGSLSGYGYPIDENQRISFGLNADNTK 541			
DB 476	PYFTVDGVSILGGLFYNDFQADDADLSDTNKSYGVDVLGPFIPNEYNSLRAGLGVVHNS 535			
QY 542	LHGGRFWMGISNNKQMA-----DGCKIQVDNNGIDPFKHDTYTNAILGWNYSLSLRPV 595			
DB 536	L-----SNMQPQAMRWLYLSMGEHPSSTDQNSFKTDDFTFN--YGVWYNKLDUR 585			
QY 596	FPTQGG--MHSVDLTVFGFDKTHQKVYVYOGNIYRPF--IKKSVLRGYAKLGYGNL---- 647			
DB 586	FPTDGSRVNLTKGVITPGSDNEYKYKVLDTATVTPIDDDHKWVLGTRRWGYDGLGGKE 645			
QY 648	PFYENFYAGGYSVRGYDOSSLGPRS-----QA-----YLTARRGQQTFL---GEV 690			
DB 646	MPYENFYAGGSVTRVGFSNTIGPKRAVYFPHQASNYDPDYDECATODGAKDLCKSDDA 705			
QY 691	VGNGALATFGESELLPLPKFGD--WIDQVRPVFIIEGGQVDFDTGMDKQITDLTQFKDQCA 749			
DB 706	VGNMAVASLEFITTPTFISDKYANSVRTSFWDMGTVWDTNWDSQYSGYDPTSDP--- 763			
QY 750	TABONAKAANRPILLTQDKQLRYSAGVAGATWYTPIGPLSISYAKPLNKKQNDQTDVTQFOI 809			
DB 764	-----SNIRMSAGTALOWMSPLGPVFSYAQPFKYDGDRAEQQFOFIN 806			

RESULT	5
C90631	
hypothetical protein ECs0179	[imported] - Escherichia coli (strain O157:H7, substrain



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 10:07:32 ; Search time 61.4 seconds  
(without alignments)  
1470.733 Million cell updates/sec

Title: US-09-701-711-2  
Perfect score: 813  
Sequence: 1 MRNSYKGFQVSAMTMAVM.....LNKKQNDQTDTVQFQIGSVF 813

Scoring table:  
Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Word size : 20

Total number of hits satisfying chosen parameters: 2

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : A\_Geneseq\_032802.\*  
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	813	100.0	813 21	AA44390
2	749	92.1	813 21	AA44391

ALIGNMENTS

RESULT 1  
ID AAY44390  
ID AAY44390 standard; Protein; 813 AA.  
XX

AC AAY44390;  
XX 14-MAR-2000 (first entry)  
XX  
XX M. catarrhalis BAS027 polypeptide #1.  
XX  
XX BAS027; OMP85; outer membrane protein; otitis media; treatment;  
KW diagnosis; bacterial infection.  
XX  
XX Moraxella catarrhalis.  
XX  
XX W09963093-A2.  
XX 09-DEC-1999.  
XX 31-MAY-1999; . 99WO-EP03822.  
XX  
XX 03-JUN-1998; 98GB-0011945.  
XX 08-MAR-1999; 99GB-0005304.  
XX  
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
XX  
XX Vinals-Bassols C;  
XX  
XX WPI; 2000-105700/09.  
XX N-PSDB; AAZ29550.  
XX  
XX Novel BAS027 polynucleotide and polypeptides from Moraxella  
PT catarrhalis useful for treating M. catarrhalis infection such as otitis  
PT media -  
XX  
XX Claim 1; Fig 3; 109pp; English.  
XX  
XX The present sequence is BAS027 polypeptide, which shows significant  
CC homology to Neisseria meningitidis OMP85 outer membrane protein. It is  
CC encoded by DNA obtained from chromosomal DNA library of Moraxella  
CC catarrhalis strain MC2931 (ATCC 43617). BAS027 polynucleotide and  
CC polypeptide can be used for diagnosis and staging of disease, determining  
CC susceptibility to a disease and to prepare medicaments for treating M.  
CC catarrhalis infections, especially otitis media. The BAS027 DNA can be  
CC used as probe for screening of genetic mutations, serotype, taxonomic  
CC classification or identification. BAS027 agonists, antagonists and  
CC antibodies may be used to prevent and/or treat bacterial infections.  
XX  
XX Sequence 813 AA;  
SQ  
Query Match 100.0%; Score 813; DB 21; Length 813;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 813; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MRNSYKGFQVSAMTMAVMVNSTHAQAADFAMNDITITGLQRTVIESLQSVLPRLGQV 60  
Db 1 mrnsykgfqsamtmavmvmnsthaqaadfmndititglqrvtieslqsvlpirlgqv 60  
Qy 61 VSENLADGVKALYATGNFSDVQVYHQEGRIYQVTERPLIAEINFEGRNRLPKEGQLQEG 120  
Db 61 vsenladgvkalyatgnfsvdqvyhqegriyqvtterpliaelnfeqnrllpkeglqeg 120  
Qy 121 LKNAGLAVGQPLKATVOMIETELTNQYISOGYYNTEITVTKOTMLDGNRKLDMTFAGK 180  
Db 121 lknaglavgqplkqatvgmieltnqyisggyynteitvktqmidgnrvklidmtfaegk 180  
Qy 181 PARVVDINIGNHFSADLLDVLAIKDNKINPLSKADRYQTEKLVTSLENLRAKYLNAG 240  
Db 181 parvvdinignhfsadlldvlaiakdnkinplskadrytqektlvtslenlrakyl nag 240  
Qy 241 FVRFEIKDAKLNINEDKNRIFVEISLHEGEQYRFQGTQFGLNLTYTQAELEALLFKAAE 300  
Db 241 fvrfeikdaaklninedknrifeislhgeqyrfqgtqfqlnltytqaeleallfkkee 300  
Qy 301 GFSQAMLEQTTNNISTKFGDDGYAAQIRPVTRINDESRSTDVVEYIDFVHPVYRRINF 360  
Db 301 gfsqamleqttnnistkfgddgyaaiqvtrindesrstdvveyidfvhpyvrrinf 360

Db 301 gfsqamleqtnnistkfgddgyvyaqirpvttrindesrtvdveyidvphvpyvrrinf 360  
QY 361 TGNFKTQDEVLRREMQLLEGALASNOKIQLSARLMRTGFFKHVTVDTTRVPNPSPOVDV 420  
Db 361 tgnfktqdevlrrremrqllegalasnkgqlsrlarlmrtgffkhvtdtrvpnpdpqdv 420  
QY 421 NFVVEQPSGSSSTIAAGYSQSGGVTFQFVSONNFMTGKHVNASFSRSSTRETVYSLGMT 480  
Db 421 nfveeqpsgssstiaagysqsggvttfqfvsqnnfmgtkghvnasfsrsetrevyslmt 480  
QY 481 NPYFTVNGVSQSLSGYRRTKYDNKNISNYVLDYSGGSLSYGYPIDENORISFGLNADNT 540  
Db 481 npyftvngvsqslsgyrrtktdydnknisnyvldsyggslysgypidengrifsflnadnt 540  
QY 541 KLHGGRFMGTSNVKQLMADGGKIQVDNNGIPDFKHDTTNNAILGWNYSLSLDPVPTQG 600  
Db 541 klhggrfmgisnvkqlmadggkigvddnngipdfkhdtytnnailgwnyslsldpvpftqg 600  
QY 601 MSHSVDLTVGFGDKTHQKVYVQGYNIYRPFIKKSVLRGYAKLGYCENNLPPFYENFYAGGYGS 660  
Db 601 mshsvdltvvgfkdthqkvvyqgyniyrpfikksvlrgyakyaklgynnlpfyenfyaggygs 660  
QY 661 VRGYDQSSLGPRSQAYLTARRGQOTTLGEGVVGGNALATFGSELILPLPFKGDWIDQVRPV 720  
Db 661 vrgydgsslgprsqayltarrgqottlgevgvggnalatfgselilplpfkgdwidqvrpv 720  
QY 721 IFTEGGGOVDTTGMDKQOTIDLTFQKDPQATAEQNAKAANRPLLTQDKOLRYSGAGVATWY 780  
Db 721 iftegggovdttgmdkotidltfqkdpqataeqnakaanrplltqdkqlrysgagvatwy 780  
QY 781 TPIGPLSISYAKPLNKKONDQDTDFVQFQIGSVF 813  
Db 781 tpiGPLsisyakplnkkqndqdtvqfQigsvf 813  
RESULT 2  
AAY44391  
ID AAY44391 standard; Protein; 813 AA.  
XX  
AC AAY44391;  
XX  
DT 14-MAR-2000 (first entry)  
XX  
DE M. catarrhalis (ATCC 43617) BASB027 polypeptide #2.  
XX  
KW BASB027; OMP85; outer membrane protein; otitis media; treatment;  
KW diagnosis; bacterial infection.  
XX  
OS Moraxella catarrhalis.  
XX  
PN W09963093-A2.  
XX  
PD 09-DEC-1999.  
XX  
PF 31-MAY-1999; 99WO-EP03822.  
XX  
PR 03-JUN-1998; 98GB-0011945.  
XX  
PR 08-MAR-1999; 99GB-0005304.  
XX  
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
XX  
XX Vinals-Bassols C;  
XX  
XX WPI; 2000-105700/09.  
XX  
DR N-PSDB; AAZ29551.  
XX  
XX Novel BASB027 polynucleotide and polypeptides from Moraxella  
PT catarrhalis useful for treating M. catarrhalis infection such as otitis  
PT media  
XX  
PS Claim 1; Page 102-104; 109pp; English.  
XX  
XX The present sequence is BASB027 polypeptide, which shows significant

CC homology to Neisseria meningitidis OMP85 outer membrane protein. It is  
CC encoded by DNA obtained from chromosomal DNA library of Moraxella  
CC catarrhalis strain Mc2931 (ATCC 43617). BASB027 polynucleotide and  
CC polypeptide can be used for diagnosis and staging of disease, determining  
CC susceptibility to a disease and to prepare medicaments for treating M.  
CC catarrhalis infections, especially otitis media. The BASB027 DNA can be  
CC used as probe for screening of genetic mutations, serotype, taxonomic  
CC classification or identification. BASB027 agonists, antagonists, and  
CC antibodies may be used to prevent and/or treat bacterial infections.

XX Sequence 813 AA;

Query Match 92.1%; Score 749; DB 21; Length 813;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 749; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 QLADGVKALYATGNFSDVQVYHQEGRIYQVTERPLAETNFEGRNLTPKELGELGKNA 124  
Db 65 qladgvkalyatgnfsdvvyhqegriyqvterplaeinfegnriltpkeglgeglkna 124  
QY 125 GLAVGQPLKOATVQMIETELTNQYISQGYNTETITVKTMLDGNRVKLDMTFAEGKPARV 184  
Db 125 glavgqplkqatvqmieteltnqyisqgyntetitvktmldgnrvkldmtfaegkparv 184  
QY 185 VDINIIGNHFSDADLIDVLAIKDNKNPLSKADRYTQEKLYTSLENLRKAYLNAGVRF 244  
Db 185 vdinignhfsadlidlavlaiakdnknplskadrytqeklytslenlrakylnagvrfr 244  
QY 245 EIKDAKLNINEDKNRIFVEISLHEGEQYRFQOTQFLGNLTVTQAELEALLFKAEQFSQ 304  
Db 245 eikdaklninedknrifeislhegeqyrfqotqfignltvtqaeleallfkaeegfsq 304  
QY 305 AMLEQTTNNISTKFGDDGYVYQAIQIRPVTRINDESTVDVEYIIDVHPVYVRRINFTGNF 364  
Db 305 amleqttnnistkfgdgyvyaqirpvttrindestvdeyviidvphvpyvrrinf 364  
QY 365 KTQDEVLRREMRLLEGALASNOKIQLSARLMRTGFFKHVTVDTTRVPNPSPOVDVNFV 424  
Db 365 ktqdevlrrremrqllegalasnkgqlsrlarlmrtgffkhvtdtrvpnpdpqdvnnfv 424  
QY 425 EEOQPSGSTIAAGYSQSGGVTFQFVSONNFMTGKHVNASFSRSSTRETVYSLGNTNPF 484  
Db 425 eeoqpsgstiaagysqsggvttfqfvsqnnfmgtkghvnasfsrsetrevyslgn 484  
QY 485 TVNGVSQSLGYRRTKYDNKNISNYVLDYSGGSLSYGYPIDENORISFGLNADTKLHG 544  
Db 485 tvngvsqslgyrrtktdydnknisnyvldsyggslysgypidendrifsflnadtkl 544  
QY 545 GFWGTSNVKQLMADGGKIQVDNNGIPDFKHDTTNNAILGWNYSLSLDPVPTQGMSSH 604  
Db 545 gfwgtsnvkqlmadggkigvddnngipdfkhdtytnnailgwnyslsldpvpftqgmshs 604  
QY 605 VDLTVGFGDKTHQKVYVQGYNIYRPFIKKSVLRGYAKLGYCENNLPPFYENFYAGGYSVRGY 664  
Db 605 vdltvfgdkthqkvvyqgyniyrpfikksvlrgyakyaklgynnlpfyenfyaggy 664  
QY 665 DQSSLGPRSQAYLTARRGQOTTLGEGVVGGNALATFGSELILPLPFKGDWIDQVRPVIFTE 724  
Db 665 dqsslgprsqayltarrgqottlgevgvggnalatfgselilplpfkgdwidqvrpv 724  
QY 725 GGOVFDTTGMDKQOTIDLTFQKDPQATAEQNAKAANRPLLTQDKOLRYSGAGVATWYPTIG 784  
Db 725 ggovfdttgmdkotidltfqkdpqataeqnakaanrplltqdkqlrysgagvatw 784  
QY 785 PLISISYAKPLNKKONDQDTDFVQFQIGSVF 813  
Db 785 plisisyakplnkkqndqdtvqfQigsvf 813

---





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 10:09:42 ; Search time 25.27 Seconds  
(without alignments)  
785.833 Million cell updates/sec

Title: US-09-701-711-2  
Perfect score: 813  
Sequence: 1 MRNSYFKGFQVSAMTMAVMM.....LNKKONDQDTVQFOIGSYF 813

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 231628 seqs, 24425594 residues

Word size : 20

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query				
No.	Score	Match	Length	DB	ID
Description					

No matches found

Search completed: September 5, 2002, 10:14:27  
Job time: 285 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 10:11:12 ; Search time 39.83 Seconds  
(without alignments)  
1961.351 Million cell updates/sec

Title: US-09-701-711-2  
Perfect score: 813  
Sequence: 1 MRNSYFKGFQVSAMTMAVM.....LNKKQNDQDTVQFQIGSVF 813

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size : 20

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : PIR\_71.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
-----						

No matches found

Search completed: September 5, 2002, 10:15:14  
Job time: 242 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 10:14:33 ; Search time 20.99 Seconds  
(without alignments)  
1499.714 Million cell updates/sec

Title: US-09-701-711-2  
Perfect score: 813  
Sequence: 1 MRNSYFKGFOVSAMTAVMM.....LNKKQNDQDTVQFIGSVF 813

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Word size : 20

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
-----						

No matches found

Search completed: September 5, 2002, 10:22:07  
Job time: 454 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 10:14:03 ; Search time 63.52 Seconds  
(without alignments)  
2214.183 Million cell updates/sec

Title: US-09-701-711-2  
Perfect score: 813  
Sequence: 1 MRNSYKGFQVSAMTMYMM.....LNKKONDQDQTVQFQIGSVF 813

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size : 20

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

- Database : SPTREMBL\_l9:\*
- 1: sp-archaea:\*
  - 2: sp-bacteria:\*
  - 3: sp-fungi:\*
  - 4: sp-human:\*
  - 5: sp-invertebrate:\*
  - 6: sp-mammal:\*
  - 7: sp-mhc:\*
  - 8: sp-organelle:\*
  - 9: sp-phage:\*
  - 10: sp-plant:\*
  - 11: sp-rodent:\*
  - 12: sp-virus:\*
  - 13: sp-vertebrate:\*
  - 14: sp-unclassified:\*
  - 15: sp-rvirus:\*
  - 16: sp-bacteriap:\*
  - 17: sp-archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
-----					

No matches found

Search completed: September 5, 2002, 10:21:40  
Job time: 457 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 09:46:41 ; Search time 88.9 Seconds  
(without alignments)  
1015.782 Million cell updates/sec

Title: US-09-701-711-2  
Perfect score: 4202  
Sequence: 1 MRNSYFKGFQVSAMTMMVMM.....LNKKQNDQDTTVOFGISVF 813

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_032802.\*  
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	4202	100.0	813	21	AA1980
2	4189	99.7	813	21	AA1981
3	1123.5	26.7	792	22	AAU03958
4	1123.5	26.7	792	22	AAU03959
5	1123.5	26.7	792	22	AAU03957
6	1115	26.5	797	22	AAU03957
7	1115	26.5	797	22	AAU03957
8	1115	26.5	797	22	AAU03957
9	1111	26.4	797	22	AAU03959
10	1111	26.4	797	22	AAU0451
11	1111	26.4	797	22	AAU0451

12	1111	26.4	797	22	AAU03958
13	1109	26.4	797	21	AAU03959
14	1104.5	26.3	792	21	AAU03958
15	1052.5	25.0	797	15	AAU03958
16	1051.5	25.0	797	15	AAU03958
17	1050.5	25.0	797	15	AAU03958
18	1047.5	24.9	793	15	AAU03958
19	1042.5	24.8	792	15	AAU03958
20	901	21.4	639	22	AAU03958
21	569	13.5	896	22	AAU03958
22	434	10.3	916	22	AAU03958
23	433	10.3	847	19	AAU03958
24	429	10.2	925	18	AAU03958
25	429	10.2	925	20	AAU03958
26	395	9.4	795	20	AAU03958
27	393	9.4	787	21	AAU03958
28	386.5	9.2	797	20	AAU03958
29	379.5	9.0	899	20	AAU03958
30	378.5	9.0	891	20	AAU03958
31	356.5	8.5	837	20	AAU03958
32	348.5	8.3	812	20	AAU03958
33	344.5	8.2	614	18	AAU03958
34	311	7.4	1249	22	AAU03958
35	256.5	6.1	578	21	AAU03958
36	249.5	5.9	576	21	AAU03958
37	198.5	4.7	469	21	AAU03958
38	197.5	4.7	469	21	AAU03958
39	197.5	4.7	469	22	AAU03958
40	180.5	4.3	624	21	AAU03958
41	180.5	4.3	643	21	AAU03958
42	180.5	4.3	818	21	AAU03958
43	176.5	4.2	919	21	AAU03958
44	176.5	4.2	1092	19	AAU03958
45	175.5	4.2	889	21	AAU03958

ALIGNMENTS

RESULT 1  
ID AAU03958 standard; Protein; 813 AA.  
AC AAU03958  
DT 14-MAR-2000 (first entry)  
DE M. catarrhalis BAS027 polypeptide #1.  
KW BAS027; OMP85; outer membrane protein; otitis media; treatment;  
KW diagnosis; bacterial infection.  
OS Moraxella catarrhalis.  
XX WO9963093-A2.  
XX 09-DEC-1999.  
XX 31-MAY-1999; 99WO-EP03822.  
XX 03-JUN-1998; 98GB-0011945.  
XX 08-MAR-1999; 99GB-0005304.  
(SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
XX Vinals-Bassols C;  
XX WPI; 2000-105700/09.  
XX N-PSDB; AAU03958.  
XX Novel BAS027 polynucleotide and polypeptides from Moraxella  
XX catarrhalis useful for treating M. catarrhalis infection such as otitis  
XX media

XX	PS	XX	Claim 1; Fig 3; 109pp; English.
CC	CC	CC	The present sequence is BASB027 polypeptide, which shows significant
CC	CC	CC	homology to Neisseria meningitidis OMP85 outer membrane protein. It is
CC	CC	CC	encoded by DNA obtained from chromosomal DNA library of Moraxella
CC	CC	CC	cattarrhalis strain Mc231 (ATCC 43617). BASB027 polynucleotide and
CC	CC	CC	polypeptide can be used for diagnosis and staging of disease, determining
CC	CC	CC	susceptibility to a disease and to prepare medicaments for treating M.
CC	CC	CC	cattarrhalis infections, especially otitis media. The BASB027 DNA can be
CC	CC	CC	used as probe for screening of genetic mutations, serotype, taxonomic
CC	CC	CC	classification or identification. BASB027 agonists, antagonists and
CC	CC	CC	antibodies may be used to prevent and/or treat bacterial infections.
XX	XX	XX	
XX	XX	XX	Sequence 813 AA;
			Query Match 100.0%; Score 4202; DB 21; Length 813;
			Best Local Similarity 100.0%; Pred: No. 5.7e-275;
			Matches 813; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1	MRNSYFKGFQVSAMTMYMVMSTHAQAADFMANDITITGLQRVIESLQSLVPLRIGOV	60
DB	1	mrnsyfkqfqsamtmyvmvmsthaqaadfmndititglqrvtieslqslvplrlgqv	60
QY	61	VSENLADGVKALYATGNFSDVQVYHGGRIIYQVTERPLIAEINFGNRLIPKSGLBQG	120
DB	61	vseqladgkvkalyatgnfsvdqvyvhggrliiyqvtterpliaeinfegnrlipksglqeg	120
QY	121	LKNAGLAVGQPLKOATVOMIETETLNQYISOGYYTETITVKOTMLDGNRVKLDMTFAEGK	180
DB	121	lknaglavgqplkqatvqmietetlcnqyisggyntetitvkqldnrvkldmftfaegk	180
QY	181	PARVVDINIGNHFSDADLIDLVAIKDNKINPLSKADRYTQEKLVTSLENLRAKYLAAG	240
DB	181	parvvdinignqhfdsadldlvaikdnkinplskadrytqeklvtslenlrakylinaag	240
QY	241	FVREIKDAKLINEDKNRIEIVESLHGEQYRFGQTOFLGNLTQAELEALLKFKAEE	300
DB	241	fvrfeikdaklinedknrieveslhgeqyrfgqtflgnltqtgaeeallkfkaee	300
QY	301	GFSQAMLEQTTNNISTKFGDDGYYYAQRVTRINDESRTVDVEYIDPVHPVYVRRINF	360
DB	301	gfsqamleqttnnistkfgdgyyyaqrvtrindesrtvdveyyidpvhpyvrrinf	360
QY	361	TGNFKTQDEVLRREMRQLEGALASNQKIQLSRARLMRTGTFKHVTVDRTPVNSPDQVDV	420
DB	361	tgnfktqdevlrrmrqlegalasnkliqlsrarlmrtgffkhvtdvdrtpvnspdqvdv	420
QY	421	NFVVEEQPSGSTIAAGYSQSGGYTFQDFDSQNNFMGTGKHVNASFSRSEFREVYSLGWT	480
DB	421	nfvveeqpsgstiaagysqsggytfqdsqnnfmgtgkhvnasfsrsefrevyslgmt	480
QY	481	NPYFTVNGVSQSLSGYYRKTKYDNKNISNYVLDSYGSLSYGYPTIDENQRIISFGLNADNT	540
DB	481	npyftvngvsqslsgyyrktkydnknisnyvldsyygsglsygyptidenqrifsglnadnt	540
QY	541	KLHGGRPMGTSNVKQLMADGGKIQVDNNGIIPDFKHDDYTYNAILGWNYSLLDRPVFPFQ	600
DB	541	klhgrpmgtsnvkqlmadggkigvdnngiipdfkhdytyynailgwnyslldrpfvftqg	600
QY	601	MSHSVDLTVGFGDKTHQKVYQGNTRYRPFKKKSVLRGVAKLGYGNLIPFENEFAGGVGS	660
DB	601	mshsvdltvfgdkthqkvvyggnlrypfkkswlrgvaklgygnlfpfenefaggygs	660
QY	661	VRGYDQSSGLSPRSQAYLTARRQQOTTIGEVVGGNALATFGSELILPLPFKGDWDQVRPV	720
DB	661	vrgydgssglprsqayltarrqqottigevvggnalatfgselilplpfkgdwdqvrpv	720
QY	721	IFTEGGQVFTTGMDKQOTIDLTFQKDPQATAEQNAKANRPILLITQDKQLRYSGAGVATWY	780
DB	721	ifleggvfttgmdkqotidltfqkdpqaetaeqnakanrpillitqdkqlrysgagvatwy	780

Db	181		parvvdlnlgnhfsdadldwlaakdnknplskadrytqeklvtslenrakylmag	240
QY	241	FVFEIKDAKLNINEOKNRIFFVEISIHGEQVRFQGTQFLGNLTYYQAELEALLFKFAEE	300	
Db	241	fvfelfkdaklninedknrlfveisihgeqyrfqgtqflgnlntytqaeleallfkfaee	300	
QY	301	GFSQAMLEQTTNNISYKFGDGGYVYQAIRPVTRINDESRVDVEYYIDVPVHPVYVRRINF	360	
Db	301	gfsqamleqttnnisckfgddgyyqaipvtrindesrtvdveyyidvpvhpvyvrrinf	360	
QY	361	TGNFKTQDEVLRREMQLSCALASNOKIOLSRARLMRTGTFXHVTVDTTPVPNSPDQVDV	420	
Db	361	tgntkqtdevlrrremrqlegalasndkqilslarlmrtgtfkhvtvdtvtpvpspdqvdv	420	
QY	421	NFVVEEQSGSSITAAAGYSOGSGVTFQFDVSONNFMGTCKGHYNASFSRSETREVYSLGMT	480	
Db	421	nfvveeqsgsstiaagysqsgvtfqfdvsqnfmgtgkhvnasfsrsetrevyslgmt	480	
QY	481	NPYFTVNGVSQSLSGYRRTKYDNKNISNYLDSYSGSLSYGYPIDENQRIISFGLNADNT	540	
Db	481	npftyvngvsqslsgyyrktkydnknisnyldsygslsygypidenqrisfglnadnt	540	
QY	541	KLHGGRFMGTSNWKQLMADGGKIQVDNNGIPDPFKHDYTTYNAILGWNYSLSLDRPVPTQG	600	
Db	541	klhggrfmglisnwkqlmadggkqlvdnnglpdpfkhdtytnaillgwnyslsldrpvptqg	600	
QY	601	MSHSVDLTVGFGDKTHQKVYVQGIYRPFIKKSVLRGYAKLGVGNLPPYENFYAGGYGS	660	
Db	601	mshsvdltvvgfkgthqkvvyqgiyrfpikksvlrgyaklgygnlppiyenfyaggygs	660	
QY	661	VRGYDQSSLGPRSOAYLTARRQQTTLGEVVGGNALATFGESELILPLPFKGDWIDQVRPV	720	
Db	661	vrgydqsslgprsqayltarrqqttlgevvggnalatfgeselilplpfkgdwidqvrpv	720	
QY	721	IFTEGGQVFTTGMDKQTTDLTFQKDPQATAGNAKAAANRPILLTQDKQLURYSAGVCAWY	780	
Db	721	ifteggqvftdtgmdkqtldltfqkdpqataeqnakaanrpilltqdkqlrysagvgatwy	780	
QY	781	TPIGPLSISYAKPLNKKQNDQDTVQFQIGSVF 813		
Db	781	tpigplsisyakplnkkqndqtdtvqfqi gsvf 813		
RESULT 3				
AAU03958				
ID	AAU03958 standard; Protein; 792 AA.			
XX	AC			
XX	AAU03958;			
DT	23-OCT-2001 (first entry)			
XX	Neisseria meningitidis serogroup A antigenic protein #1.			
DE	Serogroup B antigen; pharynx; meningitis; septicaemia; mammalian cell			
KW	bacterial infection; baculovirus; yeast.			
OS	Neisseria meningitidis.			
XX	Key Location/Qualifiers			
FH	Peptide 1..21			
FT	/note= "Signal peptide"			
FT	22..792			
FT	/note= "Mature N. meningitidis serogroup A antigen"			
XX	WO200138350-A2.			
XX	31-MAY-2001.			
XX	28-NOV-2000; 2000WO-IB01851.			
XX	29-NOV-1999; 99GB-0028197.			
PR	09-MAR-2000; 2000GB-0005698.			
PR				

	(CHIR-) CHIRON SPA.			
XX	PA PA	(STAT-) STATENS INST FOLKEHELSE.		
XX	PI Giuliani MM,	Pizza M, Rappuoli R, Holst J;		
XX	DR WPT; 2001-381289/40..			
XX	N-PSDB; AAS07278.			
XX	Novel 85 kDa antigen from Neisseria meningitidis and Neisseria gonorrhoeae, useful in the manufacture of a medicament for treating and preventing Neisserial bacteria infection -			
XX	Claim 1; Page 66-68; 92pp; English.			
CC	The sequence represents a Neisseria meningitidis serogroup A 85 kDa antigenic protein. Neisseria meningitidis colonises the pharynx, causing meningitis and, occasionally, septicaemia in the absence of meningitis. This antigenic protein is useful in the manufacture of a medicament for treating or preventing infection due to Neisseria bacteria, such as meningitis and septicaemia. It is also useful as a diagnostic reagent for detecting the presence of Neisseria bacteria or antibodies raised against Neisseria, and as a reagent for raising the antibodies. The Neisserial nucleotide sequences can be expressed in a variety of different expression systems, for example, mammalian cells, baculoviruses, plants, bacteria and yeast.			
CC	Note: There are two versions of this sequence displayed in the specification (see AAU04451).			
XX	SQ Sequence	792 AA;		
	Query Match	26.7%; Score 1123.5; DB 22; Length 792;		
	Best Local Similarity	32.6% Preq No. 2.3e-67;		
	Matches	273; Conservative 153; Mismatches 332; Indels 79; Gaps 19;		
Qy	10 QVSAMTMAVMMVMSTHAQAADFMANDITITGLQRVTIESLSVLSPFRLGVVVSENQLADG	69		
Dd	: : : l:l:l :	l:l:l :l:l:l :l:l:l :	:	:
Dd	2 klkqiassalmglspisafadftiqdirveqlgstpstvnylpkvvgdyntghsaI	61		
Qy	70 VKALYATGNFSDOVYHQEGRIIVQTERTPLIAINEFGNLPILPKESGLEKLNAGLAVG	129		
Dd	:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:	:	:	:
Dd	62 ikslyatgffdvrvetadatdgllltvtierptlgslnitgakmlndaikknesfglaqs	121		
Qy	130 QPLKAOTQMIELTINQYSIGSQYNFTVKQPMLDGNRVKLDMTEAEGKPAPVDINI	189		
Dd	:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:	:	:	:
Dd	122 qyfnoqtlnqavaglkkeylgrgklnglqipktvkcklarrrvdiditidegksakitlef	181		
Qy	190 IGNOHFSDAIDLIDLAIKDNKI-NPLSKADRYTOEKLVTSLENIARAKYLNAGFVFREKD	248		
Dd	l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:	:	:	:
Dd	182 egncqvysdrklmrqmstieggwtwlrtsrdfrfktaqdmekevctdfynngyfdfrild	241		
Qy	249 AKLINEDKNRIFVEISLHEGEQRFQCOTFLGNLT-YTONELBALKLFRAEFGFSAML	307		
Dd	:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:	:	:	:
Dd	242 tdiqtnekdrtqrkitvhheggrfwkgvsiegdtnevpkaelkkiltmtpgkwyrqqm	301		
Qy	308 EQTTNNISTFKGDGGYYAAOIRPVRTINDSRTVDVEYIDPVHPVVRINFNTGNFKTQ	367		
Dd	:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:	:	:	:
Dd	302 tavlgelnrmgsaggayselsvglpnagnaktcdfvlhiepgtkryivnelhtcgnnktr	361		
Qy	368 DEVLRRMRQLEGALASNQIKLSRARLMRTGFFKHVTVTDTRVPVPNSPDQDVNFVVEEQ	427		
Dd	l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:	:	:	:
Dd	362 devvrrelrmesapytssklrqskervellgyfdnvqfdavplagtcdkvdnlmsiter	421		
Qy	428 PSGSSTTAAGYSQGCVTFDFDVSONFMGTGKKHVNASFRSETPREVSYLGMTNPYTWN	487		
Dd	:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:	:	:	:
Dd	422 stgsldlsagvwgtglmvaagsvsgdnlfgtgasaalrasrstkltnlngsisftdpfyfad	481		
Qy	488 GVSSQSLSGY--YRK-----KYDNKNISNVLDSCGSLSYGYPIDENORISFGLNADNTKL	542		
Dd	l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:l:	:	:	:
Dd	482 gvs---lgydiygfkafpdkastsvskytttagtgvmrgipvtekydrvnflgaehlv	538		
Qy	543 HG-----GRPMGISNVKOLMGADGGKIOVDNNGIPDPFDHKDDTYTNAILGW	587		

```

Db      539 ntynkapkryadfirkygktdg-----adg-----sfkg--llykgtvgwg 577
QY      588 YSSLDPRVPTQGMHSVDLTGVF-GDK-THQKVWYVGGNIYRPFPIKKSVLRL-----GYAK 640
Db      578 rnktdsaswptrgyltgvnaeialpgsklqyysathnqtfflpskftlmlggevagiag 637
QY      641 LGYG--NNLPFYENFYAGGYSGVGYDQSSLGPRSQAYLTARRQQQTTLGDEVV--GGNAL 696
Db      638 -gygrtkelpfenfyggglgsvrgyesgtlpgk-----vydeygekisyggnkk 686
QY      697 ATFGSELILPLPKGWIDQVRPVIFFEGGVFDFTGMDKQTTIDLTQFKDPQATAEQNAK 756
Db      687 anvsaellfmpgakkd-artvrlsifadagvsw-----dgrtytaeangnnksvysenah 740
QY      757 AANRPLLTQDKQLRYSAGVGTATPTIGPLSISYAKPLNKKQNDQTDVTQVQIGSVF 813
Db      741 ks-----tftnelrysagvavtwsplgpmkfsyayplkkkpedelqrfqglgttf 792

RESULT 4
AAB84745
ID      AAB84745 standard; Protein; 792 AA.
XX      AC      AAB84745;
XX      DT      17-SEP-2001 (first entry)
XX      DE      Amino acid sequence of a Neisseria gonorrhoeae protein.
XX      KW      Serogroup B protein; outer membrane protein; Neisseria infection;
XX      KV      vaccine.
XX      OS      Neisseria gonorrhoeae.
XX      FH      Key      Location/Qualifiers
XX      FT      Peptide      1..21
XX      FT      /note= "signal peptide"
XX      FT      Protein      22..792
XX      FT      /note= "mature protein"
XX      PN      WO200152885-A1.
XX      PD      26-JUL-2001.
XX      PF      17-JAN-2001; 2001WO-IB00166.
XX      PR      17-JAN-2000; 2000GB-0001067.
XX      PR      09-MAR-2000; 2000GB-0005699.
XX      PA      (CHIR-) CHIRON SPA.
XX      PI      Pizza M, Rappuoli R, Giuliani M;
XX      WPI: 2001-451895/48.
XX      N-PSDB: AAH42129.
XX      Composition for treating or preventing infection to, detecting, or for
XX      raising antibodies against Neisseria bacteria, comprises an N.
XX      meningitidis serogroup B outer membrane preparation and an immunogenic
XX      component -
XX      Disclosure; Page 65-67; 83pp; English.
XX      The present sequence represents a Neisseria gonorrhoeae protein. The
XX      protein is used to produce the compositions of the invention. The
XX      specification describes a composition, comprising a Neisseria
XX      meningitidis serogroup B outer membrane preparation and an immunogenic
XX      component. The immunogenic component is protein disclosed in WO99/57280,
XX      WO99/36544, WO99/24578, WO99/66791, WO97/28273, WO96/29412, WO95/03413,
XX      WO99/31132, WO99/58683, WO99/55873, and/or N. meningitidis protein PorA,
XX      TbpA, TbpB, PilC, OpA, or Omp85. The composition is used for making
XX      a medicament for treating or preventing infection due to Neisseria

```

```

CC      bacteria; a diagnostic reagent for detecting the presence of Neisseria
CC      bacteria or of antibodies raised against Neisseria bacteria; and/or
CC      a reagent which can raise antibodies against Neisseria bacteria. It may
CC      also be used as a vaccine.
XX
SQ      Sequence      792 AA;

Query Match      26.7%; Score 1123.5; DB 22; Length 792;
Best Local Similarity 32.6%; Pred. No. 2.3e-67;
Matches 273; Conservative 153; Mismatches 332; Indels 79; Gaps 19;

QY      10 QVSAMTMAVMVMVSTHAQAADPFAMANDITITGLQRTVTSLSQSVLPFRGLQGVVSENQLADG 69
Db      2 kllqiasalmlglisplafadftlqdrveglrtepstvfnylpvkvgdytdnthdsai 61
QY      70 VKALYATGNFSDQVYHQEGRRIYQVTERPLIAEINFEGRNLIPKEGLQEQLKKNAGLAVG 129
Db      62 iklyatagfddrvvetadgqllltvierptgslntgkmlqndaklnlesfglaqs 121
QY      130 QPLKQATVQMIEETELTNQYISQGYNTETITVKOTMLDGNRVKLTDMTFAEGKPARVDINI 189
Db      122 qyfnqatlnqavaglkkeylgrgklniqitpkvtklarnrvdiditidegksakitdlf 181
QY      190 IGNOHFSADLIDVIAIKDNKI-NPLSKADRYTOEKLVTSLLENRAKYLNAGFVRETKD 248
Db      182 egnvqysdrklmrqmsltgegiwtlrsdrfrqkfaqdmekvtdfyngngyfdfrild 241
QY      249 AKLINEDKNRIFVEISLHGEQYRFQTOFLGNLT-YTQAELEALLKFAEEGFSQAML 307
Db      242 tdiqtnedktrtkitvhggrfrfvkvslegdnevpkaeleklmtkpgkyeyeqqm 301
QY      308 EQTTNNISTKFGDDGYIAQIRPVTRINDESRVTDVEYYIDPVHPVYVRRINFTGNFTQ 367
Db      302 tavlgeiqnrmgsagayseisvqplpnagktkdvfhlhepgrkiyveihitgnnkr 361
QY      368 DEVLRRERMOLEGALASNQIKQLSRALMRTEGFKHVTDRPVPNSDPDQDVNFVBEQ 427
Db      362 devvreirqlmesapydtsklqrskervellgyfdnvqfadvplagtpdkvdlmsiter 421
QY      428 PGSSTIAAGYSQSGGVTFQFDVSONNPMFGTKHVNASFSESTREYVSLQMTNPYFTVN 487
Db      422 stgslidsagwvqdtglvmsagvsqdnlfgtgksaalrasrsktllngslsftdpyftad 481
QY      488 GVSQSLSGY--YRKT---KYDNKNISNVLDYSGLSYGYPIDENQRISEGLNADNKL 542
Db      482 gvs---lgydiygfakfprkastsvkgyktttagggvrmgipvtceydrvnfglaehltv 538
QY      543 HG-----GREMGISNVKQLMADGGKIQVDNNGIPDFKHDTYTYNAILGNW 587
Db      539 ntynkapkryadfirkygktdg-----adg-----sfkg--llykgtvgwg 577
QY      588 YSSLDPRVPTQGMHSVDLTGVF-GDK-THQKVWYVGGNIYRPFPIKKSVLRL-----GYAK 640
Db      578 rnktdsaswptrgyltgvnaeialpgsklqyysathnqtfflpskftlmlggevagiag 637
QY      641 LGYG--NNLPFYENFYAGGYSGVGYDQSSLGPRSQAYLTARRQQQTTLGDEVV--GGNAL 696
Db      638 -gygrtkelpfenfyggglgsvrgyesgtlpgk-----vydeygekisyggnkk 686
QY      697 ATFGSELILPLPKGWIDQVRPVIFFEGGVFDFTGMDKQTTIDLTQFKDPQATAEQNAK 756
Db      687 anvsaellfmpgakkd-artvrlsifadagvsw-----dgrtytaeangnnksvysenah 740
QY      757 AANRPLLTQDKQLRYSAGVGTATPTIGPLSISYAKPLNKKQNDQTDVTQVQIGSVF 813
Db      741 ks-----tftnelrysagvavtwsplgpmkfsyayplkkkpedelqrfqglgttf 792

RESULT 5
AAB23786
ID      AAB23786 standard; Protein; 792 AA.
XX

```



PI Giuliani MM, Pizza M, Rappuoli R, Holst J;  
XX WPI; 2001-381289/40.  
DR N-PSDB; AAS07277.  
XX Novel 85 kDa antigen from *Neisseria meningitidis* and *Neisseria*  
PT gonorrhoeae, useful in the manufacture of a medicament for treating and  
PT preventing *Neisseria* bacteria infection -  
XX Claim 1; Fig 1; 92pp; English.  
PS The sequence represents a *Neisseria meningitidis* serogroup B 85 kDa  
XX antigenic protein. *Neisseria meningitidis* colonises the pharynx, causing  
CC meningitis and, occasionally, septicaemia in the absence of meningitis.  
CC This antigenic protein is useful in the manufacture of a medicament for  
CC treating or preventing infection due to *Neisseria* bacteria, such as  
CC meningitis and septicaemia. It is also useful as a diagnostic reagent for  
CC detecting the presence of *Neisseria* bacteria or antibodies raised against  
CC *Neisseria*, and as a reagent for raising the antibodies. The *Neisseria*  
CC nucleotide sequences can be expressed in a variety of different  
CC expression systems, for example, mammalian cells, baculoviruses, plants,  
XX bacteria and yeast.  
SQ Sequence 797 AA;

Query Match 26.5%; Score 1115; DB 22; Length 797;  
Best Local Similarity 32.5%; Pred. No. 8.8e-67;  
Matches 266; Conservative 154; Mismatches 361; Indels 38; Gaps 14;

QY 10 QVSAMTAVMMVMGTHAQAADFMANDITITGLQRTVETESLOSFLPRLGQVVSNNQLADG 69  
Db 2 kkgiasalmmglspialadftqdirveglrtepstvfnylpkvvgdyndthgsai 61  
QY 70 VKALYATGNFSOVYHOGRRIIVQTERPLIAEINFGNRLIPKEGLQGLKKNAGLAVG 129  
Db 62 ikisiyatgfdvdrvetadgllitvterptglsinitgkmlqndgaiknlesfglaqs 121  
QY 130 OPLQATVOMIETETNOYISOGYNTETITVKOTMLDGNRVKLDMTFAEGKPARVVDINI 189  
Db 122 qyfngatqavagkeeyigrklnlqtpkvtklarnrvdiditidegksakitdief 181  
QY 190 IGNOHFSADLIDVLAIKDNKI-NPLSKADRYTOEKLVTSLIENRAKYLNAGFVRFEKD 248  
Db 182 egngqvysarklmrgmsltgglwtlwtlrsngfneqfagdmekvtdfygnngydfdrild 241  
QY 249 AKLNINDEKNRIFVETISLHEGEQVREGOTQFLNLT-VTQAELEALLKFAEEGFSQAML 307  
Db 242 tdiqtnedktktikitvhegrfrfwgkvsiegtnevpkaeleklmckpgkwyergqm 301  
QY 308 BOTTNNISTKFGDGYVYAQRPVTRINDESRVTVDVEYIIDVPVYVVRINFTGNFKTQ 367  
Db 302 tavlgelqnrngsagayseisvqlpnaektvdfvhlhegrkiyvneihitgnnktr 361  
QY 368 DEVLREMRQLEGALASNOKTQLSRARLMRTGFFKHVTVDTRPVPNSPDQVDVNVVEEQ 427  
Db 362 devvrelrqnmesapydsklqrskervellgyfndvqavplagtpdkvdlmmslter 421  
QY 428 PSGSSTIAAGYSQSOGVTFQDVSONNFMGTGKHVNASFSTRREVYSLGNTNPFYTVN 487  
Db 422 stgslidlsagvqdtgvlmsagvsqdnlfgtgksaalraskrtklngslsftdpyftad 481  
QY 488 GVSQSLSGYR--KTKYDNKNISNVLDSSYGSLSYGYPIDENQRISFGLNADNTKLHGG 545  
Db 482 gvslygdyvgkafdrprkastsikqktttagagirmsvpteydvnglvaehltvn-- 539  
QY 546 RPMGLSNVQLMADGGKIQVDNNGIPDPFKHDYTTNAILGNWYSLSLDRPVPVTOGMSHSV 605  
Db 540 ---tynkapkhyadfkkyktgdg--tdgsfkgwlykgtvgwgrnkttdsalwptrgyltgv 595  
QY 606 DLTVGF-GDK--THQKVYVQGNRYRPFIRKSVLR-----GYAKLGYG--NNLPFFYENFAG 656  
Db 596 naelialpsklqysatnqtwtfflpslktftlmkgvegiag-gygrtkelpffenfygg 654

QY 657 GYGSVRYGDOSSLGPRSOAYLTARRGQOTTLEVV--GGNALATFGSELILPLPFKGDWI 714  
Db 655 glgsrvrgyesgtlgpk-----vydeygekisvgnknkanvsaelilfmpgskd-a 703  
QY 715 DOVRPVIFIEGGQVDFDTGMDKOTIDLTQFKDPOATAPQNAKAANRPLLTODKQLRYSG 774  
Db 704 rtvlslsfadagsvwdgktyddnssat-----ggrvqniygaguthkstkftneirysag 758  
QY 775 VGATWYTPIGPLSISYAKPLNKKQNDQTDTVQFQIGSVF 813  
Db 759 gavlswlspglpmkfsyayplkkkpedeigrfqlgttf 797  
RESULT 7  
AAB84744  
ID AAB84744 standard; Protein; 797 AA.  
XX AAB84744;  
XX AC  
XX 17-SEP-2001 (first entry)  
DT Amino acid sequence of a *Neisseria* serogroup B protein.  
XX Serogroup B protein; outer membrane protein; *Neisseria* infection;  
XX vaccine.  
XX *Neisseria meningitidis*.  
OS  
FH Key Location/Qualifiers  
FT Peptide 1..21 /note= "signal peptide"  
FT Protein 22..797 /note= "mature protein"  
FT  
XX WO200152885-A1.  
PN 26-JUL-2001.  
PD  
XX 17-JAN-2001; 2001WO-IB00166.  
XX  
XX 17-JAN-2000; 2000GB-0001067.  
PR 09-MAR-2000; 2000GB-0005699.  
XX  
XX (CHIR-) CHIRON SPA.  
FA  
PI Pizza M, Rappuoli R, Giuliani M;  
XX WPI; 2001-451895/48.  
DR N-PSDB; AAH42128.  
XX Composition for treating or preventing infection to, detecting, or for  
PT raising antibodies against *Neisseria* bacteria, comprises an N.  
PT meningitidis serogroup B outer membrane preparation and an immunogenic  
PT component -  
XX  
FS Disclosure; Page 59-61; 83pp; English.  
XX The present sequence represents a *Neisseria* serogroup B protein. The  
CC protein is used to produce the compositions of the invention. The  
CC specification describes a composition, comprising a *Neisseria*  
CC meningitidis serogroup B outer membrane preparation and an immunogenic  
CC component. The immunogenic component is protein disclosed in WO99/57280,  
CC WO99/36544, WO99/24578, WO99/66791, WO99/28273, WO96/29412, WO95/03413,  
CC WO99/31132, WO99/58683, WO99/58773, and/or N. meningitidis protein PorA,  
CC TbpA, TbpB, PilC, Opa, or Omp85. The composition is used for making  
CC a medicament for treating or preventing infection due to *Neisseria*  
CC bacteria; a diagnostic reagent for detecting the presence of *Neisseria*  
CC bacteria or of antibodies raised against *Neisseria* bacteria; and/or  
CC a reagent which can raise antibodies against *Neisseria* bacteria. It may  
CC also be used as a vaccine.  
XX Sequence 797 AA;

Query Match	26.5%	Score 1115;	DB 22;	Length 797;
Best Local Similarity	32.5%;	Pred. No. 8.e-67;		
Matches	266;	Conservative 154;	Mismatches 361;	Indels 38;
Gaps	14;			
QY	10	QVSAMTMAVMVMSTHAQAADFANDITITGLQRTVIESLQSVLPRLQGVSENQADG	69	
DB	2	klkqiasalmmlglspialadftiqdrveglrtqetstvfnylpvkvgdyndthgsai	61	
QY	70	VKALYATGNFSDVOVYHQBGRILYQVTERPLIAEINFEGRNLIPKCEGLQGLNAGLAVG	129	
DB	62	iksllyatgfdvrvetadgqllltvterptgslntgskmlqndaknklesfglaqs	121	
QY	130	QPLKQATVQMIETELTNQYISOGYNYTEIVKOTMLDGNRVKLDMTFAEGKPARVVDINI	189	
DB	122	qyfnqatinqavagkeeylgrgklniqtpkvtklarnrvdiditdegksakitdief	181	
QY	190	IGNQHFSDADLDVLAIKDKNI-NPLSKADRYTQEKLVTSLENLRAKYLNAGVFRFEIKD	248	
DB	182	egnvysdrklmrqmsltgglwtlrsnqfneqfaqdmekvdfyqngydfdrild	241	
QY	249	AKLNINEDKNRIFVEISLHGEQYRFGQTQFLGNLT-YTQAELEALLKPKAEGFSGQAML	307	
DB	242	tdiqtnedkktqiktvtvheggfrfwgksiegdtnvpevkaeleklmtkpgkwyerqgm	301	
QY	308	EQTNNISTKFGDDGYYYAQIRPVRTRINDESRVTVDVEYIDPVHPVVRINFTGNPKTO	367	
DB	302	tavlgelqnrmsgagysayseisvqplnaeaktvdfvlhieprklyvneihitgnnkr	361	
QY	368	DEVLRREMOLEGALASNQIKLSARLMTGTFKKHVTVDTRVPNSPDQVDVNFVVEEQ	427	
DB	362	devlrrelqmesapydtsklrskervellyfdnvgfdavplagtpdkvdlmslter	421	
QY	428	PSGSSTIAAGYSOGGVTFOFDYSONNFMCTGKHVNASFSETRREYVSLGTMNPFYTN	487	
DB	422	stgslslsagvwdqgtvlmsagvsqndlftgksaalrarskrtllngslsfcdpyfad	481	
QY	488	GVSOQLSGYYR--KTKYDNKNISNYLDSYGGSLSYGYPIDENQRISFGLNADNKLHGG	545	
DB	482	gvslygdyvygkatdprkastsikqykttagaglrmsvpteydrvnfglvaehltvn--	539	
QY	546	RFWGISNVKQLMDGKIGQVDNNGIPDFKHDYTYNAILGNWYSSLDPRVPFPGGSHSV	605	
DB	540	---tynkaphyadfkkygktdgt-dtgsfkwykgtvgwgrnktksalwptrgyltg	595	
QY	606	DLTVGFG-GDK-THQKVYQGNIRVPFTFKSVLR-----GYAKLGYG--NNLPFVNFYAG	656	
DB	596	naeialpgsklqysathnqtwffpiskttflmlggevgiag-gygrtkaipfenfygg	654	
QY	657	GYGSRVGYDOSSIGPRSQAYLTARRGOQTTLGVEV--GGNALATFGSELILPLPFKGDWI	714	
DB	655	glgsrvgyesgtlpgk-----vyydeygekisyggnkkanvsaelifpmpgagd-a	703	
QY	715	DQVRPVFIIEGQVDFDTGMDKOTIDLTFQKDPQATAEQNAKANRPLLTQDKQLRYSAG	774	
DB	704	rtvlslfaagsvwdgktyddnssat-----grrvqniyggagnthksftnelryrsag	758	
QY	775	VGATWTPTIGPLSISYAKPLNKKQNDQDTVPQIGSVF	813	
DB	759	gavtlwslpgpmkfsyayplkkkpedeigrfqglgttf	797	
RESULT	8			
AAB23784				
ID	AAB23784	standard; Protein; 797 AA.		
XX	AAB23784;			
AC	AAB23784;			
XX				
DT	12-JAN-2001	(first entry)		
XX				
DE	Neisseria meningitidis serogroup B amino acid sequence.			
XX				



```
QY 368 DEVLREMRQLEGALASNOKIQLSRARLMRTGFFKHVTVDTRPVNSPDQDVNFVVEEQ 427
Db 362 devvrelrimesapydtskqrskervellgyfdnvqfdaavplagtpdkvdlmnsalter 421
QY 428 PSGSSTIAAGYSQSGVTFQFDVSONNFMGTGKHVNASFSSRSETREYVYSLGMTNPVFTVN 487
Db 422 stgslidsagwvqdtglvmsagvsqdnlfgtksaalrasrskttlingslsfddpyftad 481
QY 488 GVSQSLSGYR--KTKYDNKNISNVLDSTGGSLSYGPIDENQRFSGFLNADNTKLHGG 545
Db 482 gvslygydygkafdrkastsikqykttagaglrmsvpteydrvnfgivaehltn-- 539
QY 546 RFGMSNVKQLMADGGKIQVDNNGIPDFKHDTYTNAILGNWYSSLDPRVPFPQGMHSV 605
Db 540 ---tynkaphyadfkkyktdg--tdgsfkgwlykgtvgwgrnktdsalwptrgyltg 595
QY 606 DLTVGF-GDK--THQKVYVYQGNIRYRPFIKKSVLRL-----GYAKLGYG--NNLPFENFYAG 656
Db 596 naeialpgsklqyysathnqtwffplsktftlmllggevgiag-gygrtkelpffenyf 654
QY 657 GYGSVRGYDQSSILGPRSOAYLTARRGOQFTLGEV--GGNALATFGSELILPLPFKGDWI 714
Db 655 glgsrvgyesgtlpgk-----vydeygekiysygnknkanvsaelifpmpgakt-a 703
QY 715 DOVRPVIFTEGGQVFTTGMQKQITDLTQFKDPOATAEQNAKAAANRPLLTQDKOLRYSAG 774
Db 704 rtrvlsifadagswdtktyddnssat----ggrvqniygagthkstkftnelrysag 758
QY 775 VGATWYTPIGPLSISYAKPLNKKNDQDFTVQFQIGSVF 813
Db 759 gavtclwslpgmksfayavplkpkpedelqrfqqlgttf 797

RESULT 9
AAU03959
ID AAU03959 standard; Protein; 797 AA.
AC AAU03959;
XX
XX
XX
DT 23-OCT-2001 (first entry)
DE Neisseria gonorrhoeae antigenic protein.
KW Antigenic protein; pharynx; meningitis; septicaemia; mammalian cell;
KW bacterial infection; baculovirus; yeast; Neisseria meningitidis.
OS Neisseria gonorrhoeae.
FH Key Location/Qualifiers
FT Peptide 1..21
FT Protein /note= "Signal peptide"
FT Protein 22..797
FT Protein /note= "Mature N. gonorrhoeae antigen"
XX
XX WO200138350-A2.
PN
XX
XX 31-MAY-2001.
PD
XX
XX 28-NOV-2000; 2000WO-IB01851.
PF
XX
XX 29-NOV-1999; 99GB-0028197.
PR
XX 09-MAR-2000; 2000GB-0005698.
PR
XX
XX (CHIR-) CHIRON SPA.
PA (STAT-) STATENS INST FOLKEHELSE.
XX
XX Giuliani MM, Pizza M, Rappuoli R, Holst J;
XX
XX WPI; 2001-381289/40.
DR
XX N-PSDB; AAS07279.
XX
XX Novel 85 kDa antigen from Neisseria meningitidis and Neisseria
XX gonorrhoeae, useful in the manufacture of a medicament for treating and
PT
```

```
PT preventing Neisserial bacteria infection -
XX Claim 1; Page 37-39; 92pp; English.
XX
CC The sequence represents a Neisseria gonorrhoeae 85 kDa antigenic protein.
CC N. gonorrhoeae is closely related to N. meningitidis, which colonises the
CC pharynx, causing meningitis and, occasionally, septicaemia in the absence
CC of meningitis. This antigenic protein is useful in the manufacture of a
CC medicament for treating or preventing infection due to Neisseria
CC bacteria, such as meningitis and septicaemia. It is also useful as a
CC diagnostic reagent for detecting the presence of Neisseria bacteria or
CC antibodies raised against Neisseria, and as a reagent for raising the
CC antibodies. The Neisserial nucleotide sequences can be expressed in a
CC variety of different expression systems, for example, mammalian cells,
CC baculoviruses, plants, bacteria and yeast.
XX
SQ Sequence 797 AA;
Query Match 26.4%; Score 1111; DB 22; Length 797;
Best Local Similarity 32.4%; Pred. No. 1.6e-66;
Matches 265; Conservative 155; Mismatches 361; Indels 38; Gaps 14;
QY 10 QVSAMTAVVMVMSVTHQAADFMANDITITGLQRTVIESLQSVLPFRIGVQVVSNNOLADG 69
Db 2 kklqiasalmvlgisplaladftiqdlrvqglrtepstcvfnylpkvvgdyndthgsai 61
QY 70 VKALYATGNFSDVQVYHGEGRYIYQVTERPLIAETIEGNEGRNLIPIKEGLOEGLKNAGLAVG 129
Db 62 iklsiyatgfdvdrvetadgqllltvierptgslngakmlqndaknlesfglaqs 121
QY 130 QPLQOATVOMIETELTNQYISQGYNTIEITVKQTMLDGNRVKLDMTFAEGKPARVVVDNI 189
Db 122 qyfnqatlnqavaglkceylgrglniqitpkvtklarvrviditidgksaktldlef 181
QY 190 IGNQHFSDADLIDVLAIKDNKI-NPLSKADRYTOEKLVTLSLENRAKYLINAGFVFEIKD 248
Db 182 egngvysdrklnrqmsltcggilwtlrsnqfneqkfaqdmekvtdfygnngyfdfrild 241
QY 249 AKLNINEDKNRIFVEISLHEGEYRFGQTFQGNLT-YTQAELEALLFKAEEGSQAML 307
Db 242 tdiqtnedktqtkitvhggrfrwkvsviegdtnvypkacelkilmkpgkwyrqgm 301
QY 308 EOTTNNISTKFGDDGYVYVAQIRPVTRINDESTVDVEYVIDVHPVYVRRINFTGNFKTQ 367
Db 302 tavlgelqnrmsgayayseisvqplpnaekttvfvhiepgrkiyvinehitgnktr 361
QY 368 DEVLREMRQLEGALASNOKIQLSRARLMRTGFFKHVTVDTRPVNSPDQDVNFVVEEQ 427
Db 362 devvrelrimesapydtskqrskervellgyfdnvqfdaavplagtpdkvdlmnsalter 421
QY 428 PSGSSTIAAGYSQSGVTFQFDVSONNFMGTGKHVNASFSSRSETREYVYSLGMTNPVFTVN 487
Db 422 stgslidsagwvqdtglvmsagvsqdnlfgtksaalrasrskttlingslsfddpyftad 481
QY 488 GVSQSLSGYR--KTKYDNKNISNVLDSTGGSLSYGPIDENQRFSGFLNADNTKLHGG 545
Db 482 gvslygydygkafdrkastsikqykttagaglrmsvpteydrvnfgivaehltn-- 539
QY 546 RFGMSNVKQLMADGGKIQVDNNGIPDFKHDTYTNAILGNWYSSLDPRVPFPQGMHSV 605
Db 540 ---tynkaphyadfkkyktdg--tdgsfkgwlykgtvgwgrnktdsalwptrgyltg 595
QY 606 DLTVGF-GDK--THQKVYVYQGNIRYRPFIKKSVLRL-----GYAKLGYG--NNLPFENFYAG 656
Db 596 naeialpgsklqyysathnqtwffplsktftlmllggevgiag-gygrtkelpffenyf 654
QY 657 GYGSVRGYDQSSILGPRSOAYLTARRGOQFTLGEV--GGNALATFGSELILPLPFKGDWI 714
Db 655 glgsrvgyesgtlpgk-----vydeygekiysygnknkanvsaelifpmpgakt-a 703
QY 715 DOVRPVIFTEGGQVFTTGMQKQITDLTQFKDPOATAEQNAKAAANRPLLTQDKOLRYSAG 774
Db 704 rtrvlsifadagswdtktyddnssat----ggrvqniygagthkstkftnelrysag 758
```



Dd      704    rtvrlslfadagsvwdgktyddnssst-----ggrrvqnlyaganthkstfnelrysaq    758

QY      775    VGATWTPIGPLSISYAKPLNKKONDQTDTVQFOIGSVF    813  
               || :|||: ||| || | | :    ||| : |  
Dd      759    gavtwisplgmksyayplkkpedeigrfglgttf    797

RESULT 10  
AAU04451  
ID    AAU04451 standard; Protein; 797 AA.  
XX  
AC    AAU04451;  
XX  
DT    23-OCT-2001 (first entry)  
XX  
DE    Neisseria meningitidis serogroup A antigenic protein #2.  
XX  
KW    Serogroup B antigen; pharynx; meningitis; septicaemia; mammalian cell;  
XX    bacterial infection; baculovirus; yeast.  
XX  
OS    Neisseria meningitidis.  
XX  
FH    Key                  Location/Qualifiers  
FT    Peptide              1..21  
FT                      /note= "Signal peptide"  
FT    Protein             22..797  
FT                      /note= "Mature N. meningitidis serogroup A antigen"  
PN    WO200138350-A2.  
XX  
XX    31-MAY-2001.  
PD  
XX  
XX    28-NOV-2000; 2000WO-IB01851.  
PF  
XX    29-NOV-1999; 99GB-0028197.  
PR  
XX    09-MAR-2000; 2000GB-0005698.  
PR  
XX  
PA    {CHIR-) CHIRON SPA.  
PA    {STAT-) STATENS INST FOLKEHELSE.  
XX  
PI    Giuliani MM, Pizza M, Rappuoli R, Holst J;  
XX  
DR    WFI; 2001-381289/40.  
XX  
PT    Novel 85 kDa antigen from Neisseria meningitidis and Neisseria  
PT    gonorrhoeae, useful in the manufacture of a medicament for treating and  
PT    preventing Neisserial bacteria infection -  
XX  
PS    Claim 1; Page 39-40; 92pp; English.  
XX  
CC    The sequence represents a Neisseria meningitidis serogroup A 85 kDa  
CC    antigenic protein. Neisseria meningitidis colonises the pharynx, causing  
CC    meningitis and, occasionally, septicaemia in the absence of meningitis.  
CC    This antigenic protein is useful in the manufacture of a medicament for  
CC    treating or preventing infection due to Neisseria bacteria, such as  
CC    meningitis and septicaemia. It is also useful as a diagnostic reagent for  
CC    detecting the presence of Neisseria bacteria or antibodies raised against  
CC    Neisseria, and as a reagent for raising the antibodies. The Neisserial  
CC    nucleotide sequences can be expressed in a variety of different  
CC    expression systems, for example, mammalian cells, baculoviruses, plants,  
CC    bacteria and yeast.  
CC    Note: There are two versions of this sequence displayed in the  
CC    specification (see AAU03958).  
XX  
SQ    Sequence    797 AA;

Query Match                  26.4%; Score 1111; DB 22; Length 797;  
Best Local Similarity        32.4%; Pred. No. 1.6e-66;  
Matches    265; Conservative 155; Mismatches 361; Indels 38; Gaps 14;

QY      10    QVSAMTVMWMVMTSHAQAADFMANDITTCGLQRWTFIESQLPFRIGVVSENQLADG    69  
               :: : |:|:: : ||| || : |||| :: | | :: : | :: :  
               :: : |:|:: : ||| || : |||| :: | | :: : | :: :

Db	2	klkqiasalmvlgisplaladftiqdirvegqlrtepstpvfnlpykvvgdtyndthgsai	61
Qy	70	VKALYATGNFSDVQVYHQEGRIIYQTERPLIAEINFEQGNRLIPKGLQEGLNAGLAVG	129
Db	62	ikslatgfdvrvetadqgliltvterptqslnitgakmlqndaknkiesfslags	121
Qy	130	QPLKQATQVMETELNQYISQGYNYETPTVKOTMLDGNRNVKLDMTFAEKGKAPRVVDINI	189
Db	122	qyfnqatinqavaglkkeelyigrklnlqitcpkvtklarnrvdlditidegkaakidief	181
Qy	190	IGNQHSDEADLIDVLAIKDKKI-NPLSKADRYTQEKLVTSLENLRAKYLNAGVPRFEIKD	248
Db	182	egnqvysdrklmrqmsltgeglwtlrsnfneqfagdmekvtyfqnngyfdfrild	241
Qy	249	AKLINEDKNRIFVEISLHGEQYRFQOTFLGNLT-YTQAELEALLKPKAEFGFSQAML	307
Db	242	tdiqtneedtkktikittvheggrfwgkvsiegdtnevpkaeleklitmkpgkwyerqgm	301
Qy	308	EQTNNISTFGDDGYAAQIRPVTRINDESRVDVEYIIDPVHVPYVRRINPTGNFKTQ	367
Db	302	tavlgelqnmgsagyaayselsvqplpnaetkcvdlhlepgklyvneihtgnnkt	361
Qy	368	DEVLRERMRQLEGALASNQIKIOLSRARLMRTGFEKKHVTVDTRVPVNSPOVDVNFVVEQ	427
Db	362	devvrreilqmesapydtsklqrskervellgfdnvqfdavplagtdpkvdlnmslter	421
Qy	428	PSGSTIAAGYSQSGGVTFQFDYSQNNFMGTGKHVNASFSRSTREYVSLGMTNPVFTVN	487
Db	422	stgsldisagvwdgtglvmsagvsqdnlfgtgksaalrsrskttlingslsfdpyftad	481
Qy	488	GVSQSLSGYYR--KTYDYNKNIISNYVLDSYSGLSYGYPIDENQRIISFGLNADMTKLHGG	545
Db	482	gvslygdvygkafdpkastsikqyktttagaglrmsvpvteydrfnvgivaehltnv--	539
Qy	546	RFWGISNVKQLMADGGKIQVDNNGIIPDFKHIDYTYNAILGNWYSSLDPRVFTQGMHSV	605
Db	540	---tynkaphkhyadflkkygktdgt-dtgsfkglwykgtwgwgcnktdsalwptrgyltg	595
Qy	606	DLTWGF-GDK-THQKVYOGNIIVRPFKKSVLR----GYAKLGYG--NNLPYENFYAG	656
Db	596	naeialpgsklqysathnqtwfplskttflmlggevglag-gygrtkeipfenfygg	654
Qy	657	GYGSVRGYDQSSGLSPRSQAYLTARRQQTTLGEV--GGNALATFSGELTLPFPFGDWI	714
Db	655	glgsvrgyesgtlpgk-----vydeygekisygggnkkanvsaeellfmpgagd-a	703
Qy	715	DQVRPVFIISGGQVFTTGHDKQOTIDLTQFKDQPAETAQNKAANRPLLTQDKQLRYSAG	774
Db	704	rtvlslifadagsvwdgktyddnsssat-----ggvrqnlylgagnthkstkftnelrysg	758
Qy	775	VGATWTWPTPIGLSTISVAKPLNKKNQDQTDVQFOIGSVF	813
Db	759	gavtwslpigmkfsayplkkkpedeigrfqfqlgttf	797
RESULT	11		
AA	84746		
ID	AA84746	standard; Protein; 797 AA.	
XX	AA		
XX	AA		
XX	AA		
DT	DT	17-SEP-2001 (first entry)	
XX	DE	Amino acid sequence of a Neisseria serogroup A protein.	
XX	KW	Serogroup A protein; outer membrane protein; Neisserial infection;	
XX	KW	vaccine.	
XX	OS	Neisseria meningitidis.	
XX	XX		
FH	Key	Location/Qualifiers	
FT	Peptide	1..21	
FT		/note= "signal peptide"	





```
QY 249 AKLNEDKRIEVEISLHEGEYREGQTFGLNLT-YTOAELEALLKPKAEEGFSQAML 307
Db 242 tdiqnedkcktkicvheggrfrwgvksiegtnevpkaeleklitmkpgkweryqgm 301
QY 308 EQTNNISTKFGDDGYIAQIRPVTRINDSRVDVEYYIDPVHPVYVRINFTGNFKTQ 367
Db 302 tavlgeiqnrmgsagayseisvqplpnaetktvdfvlhieprklyvneihitgnnkr 361
QY 368 DEVLRRMRQLEGALASNQIKLSRARLMTGFFKHVTVDTRPVNSPDQVDVNFVVEEQ 427
Db 362 devrrrlrqlmesapydtsklrskervellgyfndvqfdaavplagtpdkvdlmslter 421
QY 428 PSGSSTIAAGYSQGGVTFQFDVSONNFMGTGKHVNASFSESRSETRVYSLCMTNPYFTVN 487
Db 422 stgslidagvwdqdtglvmsagvsqdnlfgtgksaalrasrsktllngslsfdpdyfad 481
QY 488 GVSQSLSGYR--KTKYDNKNISNYLDSYGGSLSYGYPIDENORISFGLNADTKLHGG 545
Db 482 gvslygyvgykafdrkastsikqyktttagagirmsvpvteydrvnfgivaehltn-- 539
QY 546 RFMGISNVKQLMADGGKIQVDNNGIPDFKHDTYTNAILGWNYSLSLDRPVFPPTQGMHSV 605
Db 540 ---tynkapkhyadfikygtkdg--tdgskfwlykgtvgwgrnktksalwptgryltg 595
QY 606 DLTVGF-GDK--THQKVYVQGNIRYRPFIKKSVLR-----GYAKLGYG--NNLPFYENFYAG 656
Db 596 naeialpgsklyqysathnqtwtfflskttftlmglggevgiag-gygrtkelpffenyf 654
QY 657 GYGSVRGYDQSSLSGRSOAYLTARRGQOFTLGEVV--GGNALATFGSELILPLPFKGDWI 714
Db 655 glgsrvgyesatlgpk-----vydeygekisyggnkkanvsaelifmpgpkd-a 703
QY 715 DQVRPVIIEGQVDTGMDKQITDLTQFKDQATAEQNAKANRPLLTDKOLRYSAG 774
Db 704 rtvlslsfadagsvwdktydnssat-----ggrvqnlvgagntkstkftneirysag 758
QY 775 VGATWYTPIGPLSISYAKPLNKKNDQDTVQFOIGSVF 813
Db 759 gavtlwplgpmkfrayaplkpkpedelqrfqfqlgttf 797

RESULT 14
ID AAY84946 standard; Protein; 792 AA.
XX
AC AAY84946;
XX
DT 21-AUG-2000 (first entry)
XX
DE Amino acid sequence of outer membrane protein (omp) 85.
XX
KW Outer membrane protein; omp85; gonococcal infection; symptomatic disease;
KW meningococcal infection; protective immune response; vaccine.
XX
OS Neisseria gonorrhoea.
XX
FH Key
FT Peptide Location/Qualifiers
FT /note= "signal peptide"
XX
XX WO200023595-A1.
XX
XX 27-APR-2000.
XX
XX 22-OCT-1998; 98WO-US22352.
XX
XX 22-OCT-1998; 98WO-US22352.
XX
XX (UYMO-) UNIV MONTANA.
XX
XX Judd RC, Manning SD;
XX
```

WPI; 2000-339694/29.  
N-PSDB; AAA15155.

DR New isolated outer membrane protein 85 of *Neisseria gonorrhoeae* and *N.*  
XX meningitidis useful for vaccine, therapeutic and diagnostic  
PT compositions for gonococcal or meningococcal infections  
XX  
PS Claim 1; Fig 2; 98pp; English.  
XX

CC The present sequence represents an outer membrane protein (omp) 85 of  
CC *Neisseria gonorrhoeae*. The omp polypeptides and polynucleotides are  
CC useful in compositions for use in the prevention, treatment and diagnosis  
CC of non-symptomatic gonococcal infection or meningococcal infection and  
CC symptomatic disease. They are also useful for the detection of  
CC hybridisation complexes. Antigens and antibodies specific omp proteins  
CC also provide diagnostic, therapeutic and prophylactic compositions for  
CC the treatment or prevention of the infections described above. The  
CC antibodies are useful for inducing a protective immune response in  
CC humans or animals with *N. gonorrhoeae*, *N. meningitidis*, or other  
CC *Neisseria* species. The proteins, antibodies and polynucleotide  
CC sequences of the present invention may also be used in the screening  
CC and development of chemical compounds such as drugs or vaccines.

XX Sequence 792 AA;

Query Match 26.3%; Score 1104.5; DB 21; Length 792;  
Best Local Similarity 32.6%; Pred. No. 4.4e-66;  
Matches 268; Conservative 155; Mismatches 350; Indels 49; Gaps 17;

```
QY 10 QVSAMTAVMVMVMTTHAQADFMANDITITGLQRTVIESLQSLVLPFRFGQVSVENQLADG 69
Db 2 kklqasalmmlglspafadftqdrveglrtepstvfnylpvkgvdyndchgsal 61
QY 70 VKALYATGNFSDVQVYHOEGRIIYQVTERPLIAETAEINFGNRLIPKEGLEKKNAGLAVG 129
Db 62 ikxiyatgfdvdrvetadgillltivcptgislmgakmlqndaklnlesfglags 121
QY 130 QPLKOATVOMITETLTNOYISQGYNTBITVKQTMLDGNRVKLDMTFAGGPARVDINI 189
Db 122 qyfnqatngavagikeeylgrgklniqtpkvtklarnrvdltiditdegksakltdef 181
QY 190 IGNOHFSADLDLIDVLAIKDNKI-NPLSKRADYTOEKLTSLLENRAKYLNAGVFBEIKD 248
Db 182 egnvqysdrkkmrqmslceggilwtlfrsdrfdqkfaqmekvtcdfyqngyfdfrild 241
QY 249 AKLNEDKRIEVEISLHEGEYREGQTFGLNLT-YTOAELEALLKPKAEEGFSQAML 307
Db 242 tdiqnedkcktkicvheggrfrwgvksiegtnevpkaeleklitmkpgkweryqgm 301
QY 308 EQTNNISTKFGDDGYIAQIRPVTRINDESRTVDVEYYIDPVHPVYVRINFTGNFKTQ 367
Db 302 tavlgeiqnrmgsagayseisvqplpnaetktvdfvlhieprklyvneihitgnnkr 361
QY 368 DEVLRRMRQLEGALASNQIKLSRARLMTGFFKHVTVDTRPVNSPDQVDVNFVVEEQ 427
Db 362 devrrrlrqlmesapydtsklrskervellgyfndvqfdaavplagtpdkvdlmslter 421
QY 428 PSGSSTIAAGYSQGGVTFQFDVSONNFMGTGKHVNASFSESRSETRVYSLCMTNPYFTVN 487
Db 422 stgslidagvwdqdtglvmsagvsqdnlfgtgksaalrasrsktllngslsfdpdyfad 481
QY 488 GVSQSLSGYR--KTKYDNKNISNYLDSYGGSLSYGYPIDENORISFGLNADTKLHGG 545
Db 482 gvslygyvgykafdrkastsikqyktttagagirmsvpvteydrvnfgivaehltn-- 539
QY 546 RFMGISNVKQLMADGGKIQVDNNGIPDFKHDTYTNAILGWNYSLSLDRPVFPPTQGMHSV 605
Db 540 ---tynkapkhyadfikygtkdg--tdgskfwlykgtvgwgrnktksalwptgryltg 595
QY 606 DLTVGF-GDK--THQKVYVQGNIRYRPFIKKSVLR-----GYAKLGYG--NNLPFYENFYAG 656
Db 596 naeialpgsklyqysathnqtwtfflskttftlmglggevgiag-gygrtkelpffenyf 654
QY 657 GYGSVRGYDQSSLSGRSOAYLTARRGQOFTLGEVV--GGNALATFGSELILPLPFKGDWI 714
Db 655 glgsrvgyesatlgpk-----vydeygekisyggnkkanvsaelifmpgpkd-a 703
QY 715 DQVRPVIIEGQVDTGMDKQITDLTQFKDQATAEQNAKANRPLLTDKOLRYSAG 774
Db 704 rtvlslsfadagsvwdktydnssat-----ggrvqnlvgagntkstkftneirysag 758
QY 775 VGATWYTPIGPLSISYAKPLNKKNDQDTVQFOIGSVF 813
Db 759 gavtlwplgpmkfrayaplkpkpedelqrfqfqlgttf 797
```





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 09:47:21 ; Search time 37.39 Seconds  
(without alignments)  
531.105 Million cell updates/sec

Title: US-09-701-711-2  
Perfect score: 4202  
Sequence: 1 MNSYFKGFQVSAMTAVNM.....LNKQNDQDTVQFQIGSVF 813

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1052.5	25.0	797	3	US-08-433-522A-2
2	1052.5	25.0	797	3	US-08-433-522A-4
3	1052.5	25.0	797	3	US-08-433-522A-6
4	1052.5	25.0	797	3	US-09-135-166-2
5	1052.5	25.0	797	3	US-09-135-166-4
6	1052.5	25.0	797	3	US-09-135-166-6
7	1052.5	25.0	797	3	US-08-942-046-2
8	1052.5	25.0	797	4	US-08-942-046-4
9	1052.5	25.0	797	4	US-08-942-046-6
10	1047.5	24.9	793	3	US-08-433-522A-10
11	1047.5	24.9	793	3	US-09-135-166-10
12	1047.5	24.9	793	4	US-08-942-046-10
13	1042	24.8	792	3	US-08-433-522A-8
14	1042	24.8	792	3	US-09-135-166-8
15	1042	24.8	792	4	US-08-942-046-8
16	146.5	3.5	2123	4	US-08-968-685A-10
17	142	3.4	2199	5	PCT-US95-11684-2
18	141	3.4	1599	2	US-08-617-697-9
19	138	3.3	1536	1	US-08-038-682-2
20	138	3.3	1536	1	US-08-302-832-2
21	138	3.3	1536	2	US-08-530-198-2
22	138	3.3	1536	2	US-08-469-880-2
23	138	3.3	1536	2	US-08-728-470-2
24	138	3.3	1536	2	US-08-617-697-2
25	138	3.3	1536	4	US-08-719-641-2
26	136.5	3.2	2314	4	US-09-268-347-49
27	136	3.2	990	4	US-09-627-376-7

28	134.5	3.2	1178	1	US-08-446-486-5	Sequence 5, Appli
29	134.5	3.2	1178	1	US-08-463-308-5	Sequence 5, Appli
30	134.5	3.2	1182	1	US-08-349-867-34	Sequence 34, Appl
31	134.5	3.2	1182	2	US-08-598-305A-34	Sequence 34, Appl
32	134.5	3.2	1188	1	US-08-239-476-34	Sequence 34, Appl
33	134.5	3.2	1188	2	US-08-639-923A-34	Sequence 34, Appl
34	134.5	3.2	1188	5	PCT-US95-05431-34	Sequence 34, Appl
35	134	3.2	2048	4	US-09-268-347-48	Sequence 48, Appl
36	132	3.1	905	4	US-09-074-658-70	Sequence 70, Appl
37	130	3.1	969	1	US-07-671-817A-4	Sequence 4, Appli
38	130	3.1	1177	1	US-07-828-788A-8	Sequence 8, Appli
39	130	3.1	1177	1	US-07-920-085-2	Sequence 2, Appli
40	130	3.1	1177	5	PCT-US92-11337-8	Sequence 8, Appli
41	130	3.1	1177	6	5169629-2	Patent No. 5169629
42	129.5	3.1	2089	1	US-08-418-893D-23	Sequence 23, Appl
43	129.5	3.1	2089	1	US-08-418-893D-24	Sequence 24, Appl
44	128.5	3.1	682	3	US-08-613-009A-10	Sequence 10, Appl
45	128.5	3.1	702	2	US-08-867-941-25	Sequence 25, Appl

ALIGNMENTS

RESULT 1  
US-08-433-522A-2  
: Sequence 2, Application US/08433522A  
: Patent No. 6013514  
: GENERAL INFORMATION:  
: APPLICANT: CHONG, Pele  
: APPLICANT: THOMAS, Wayne  
: APPLICANT: YANG, Yan Ping  
: APPLICANT: LOOSMORE, Sheena  
: APPLICANT: SIA, Dwo Yuan Charles  
: APPLICANT: KLEIN, Michel  
: TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN  
: NUMBER OF SEQUENCES: 55  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Sim & McBurney  
: STREET: 6TH Floor, 330 University Avenue  
: CITY: Toronto  
: STATE: Ontario  
: COUNTRY: Canada  
: ZIP: M5G 1R7  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patentin Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/433,522A  
: FILING DATE: 12-SEP-1995  
: CLASSIFICATION: 435  
: ATTORNEY/AGENT INFORMATION:  
: NAME: STEWART, Michael I  
: REGISTRATION NUMBER: 24,973  
: REFERENCE/DOCKET NUMBER: 1038-434 MIS:j.b  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (416) 595-1155  
: TELEFAX: (416) 595-1163  
: INFORMATION FOR SEQ ID NO: 2:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 797 amino acids  
: TYPE: amino acid  
: TOPOLOGY: linear  
: MOLECULE TYPE: protein  
: US-08-433-522A-2

Query Match 25.0%; Score 1052.5; DB 3; Length 797;  
Best Local Similarity 31.6%; Pred. No. 4.5e-76;  
Matches 260; Conservative 160; Mismatches 333; Indels 69; Gaps 20;  
Oy 23 STHAQADFAMNDITITGLQRTVIESLOSGLPFRLGQVVSENQLADGVKALYATGNFSDV 82

```
Db 14 TTTVFAAPFAVKDIRVDGVGDLEQOIRASIPVRACORVTDNDVANIVRSLSVSGRFDV 73
QY 83 QVYHOGRI-IYQVTERPLIAEFNFGNRLIPKEGLQGLKNAGLAVGQPLKQATVOMIE 141
Db 74 KA-HQEGDLVWSVAKSIISDVKIKGNSVITPEALKQNLNDANGFKVGYDLIREKLENEFA 132
QY 142 TELTNQYISOGYNTTEITVQKOTLDGNRVKLDMTFAEGKPARVVDINIIGNHFSADLI 201
Db 133 KSVKHYASVGRYNATPEIVNTLPNRAEILIQINEDDRKAKLASLTFKGNESVSSTLQ 192
QY 202 DVLAIKDNKINPLSKADRYTQELVTSLENLRAKYNAGVFRFEIKDAKLNINEDKNRIF 261
Db 193 EQMELQPSWKKL-WGNKFEQAQFKDQSIIRDYLYNNGYAKAQITKTDVQLNDEKTKVN 251
QY 262 VEISLHEGEQYRFGOTQFLGNLTQAELEALLK-FKAEEGFSQAMLEQTTNNISTKFGD 320
Db 252 VTIDVNEGLOYDLRSARIIGNLGMSAELEPLLSALHLNDFRSDIADVENAIKALGE 311
QY 321 DGYYIAQIRPVTINDESRTVDVEYIDPVHPVYVRINFTGNFKTQDEVLRREMRLQEG 380
Db 312 RYGSATVNSVPDFFDANKLTAITLVVDAGRRLTVRQLRFGNTVSADSTLRQEMRQEG 371
QY 381 ALASNQIKLSRARLMRTGFKHVTVDTR--PVPNSPDQVDVNVFVEQPGSGSTIAAGY 438
Db 372 TWYNSQLVELGKIRLDRGTGFFE--TVENRIDPINGSNDEVDVYKVKERNGTGSINFGIGY 429
QY 439 SOSGGVTFQFDVSONNFMCTGKHVNASFRRSETREYVSLGMTNPNYTVNGVVSQSLGYR 498
Db 430 GTESGISYQASVKQDNFLGTGAASVIACTKNDYGTSVNLGYTEPYFTKDG--SLGNGVF 487
QY 499 KTKYDNK---NISNVLDSDYSGSLSYGYPIDENORISFGL-----NADNTKLHGRFMGI 550
Db 488 FENYDNKSDTSNRYKRTYGSNLTGFPVNNNSYVGLGHTYKNKISNFALEYNRNLYI 547
QY 551 SNVKQLMADGKGIQVDNNGIPDPFKHDYTYNAILGNWYSSLDPRVFPQCGSHSVD--LT 608
Db 548 QSMK-----FKGNGIKTNDDFS-----FGWNYNSLRGYFTPKGVKASLGGGRVT 592
QY 609 VGFGDKTHOKVYVQGNIRPFE-----IKKSVLRGYAKLGYGN-NLPFFYENFVAGGYGS 660
Db 593 IPGSDNKYKLSADVOGYPFLDRHLWVWSAKASAGYAN-GFGNKRFLPYTYTAGGIGS 651
QY 661 VRGYDQSSLSGRSQAYLTFARQQOITLG-----EVVGGNALATFGSELILPLPFKGDW 713
Db 652 LRGFAYGSGIPNA---IYAEYCGNGSGTGFKKIISDVIIGNAIATASAEILVPTFPVSDK 708
QY 714 I-DQVRPIFIEGGQVETTT-GMDKQITDITQFKDPQATAEQNAKAAANRPLITQDKLRY 771
Db 709 SONTVRTSLFVDAASVWNTKWSKNGLESVDLK-----RLPDYGKSSRIRA 755
QY 772 SAGVGATWYTPIGPLSISYAKPLNKKQNDQITDVTQFQIGSVF 813
Db 756 STGVGFQWQSPIGPLFVSYAKPIKKYENDDDVQEQFQSIGSF 797
```

## RESULT 2

```
US-08-433-522A-4
; Sequence 4, Application US/08433522A
; Patent No. 6013514
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
```

```
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433,522A
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-434 MIS:jlb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 797 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-433-522A-4
```

```
Query Match 25.0%; Score 1052.5; DB 3; Length 797;
Best Local Similarity 31.6%; Pred. No. 4.5e-76;
Matches 260; Conservative 160; Mismatches 333; Indels 69; Gaps 20;

QY 23 STHAQAADFMANDITITGLQRVTIESLOSVPFRGLGVVSENQNLADGVKALYATGNFSDV 82
Db 14 TTTVFAAPFAVKDIRVDGVGDLEQOIRASIPVRACORVTDNDVANIVRSLSVSGRFDV 73
QY 83 QVYHOGRI-IYQVTERPLIAEFNFGNRLIPKEGLQGLKNAGLAVGQPLKQATVOMIE 141
Db 74 KA-HQEGDLVWSVAKSIISDVKIKGNSVITPEALKQNLNDANGFKVGYDLIREKLENEFA 132
QY 142 TELTNQYISOGYNTTEITVQKOTLDGNRVKLDMTFAEGKPARVVDINIIGNHFSADLI 201
Db 133 KSVKHYASVGRYNATPEIVNTLPNRAEILIQINEDDRKAKLASLTFKGNESVSSTLQ 192
QY 202 DVLAIKDNKINPLSKADRYTQELVTSLENLRAKYNAGVFRFEIKDAKLNINEDKNRIF 261
Db 193 EQMELQPSWKKL-WGNKFEQAQFKDQSIIRDYLYNNGYAKAQITKTDVQLNDEKTKVN 251
QY 262 VEISLHEGEQYRFGOTQFLGNLTQAELEALLK-FKAEEGFSQAMLEQTTNNISTKFGD 320
Db 252 VTIDVNEGLOYDLRSARIIGNLGMSAELEPLLSALHLNDFRSDIADVENAIKALGE 311
QY 321 DGYYIAQIRPVTINDESRTVDVEYIDPVHPVYVRINFTGNFKTQDEVLRREMRLQEG 380
Db 312 RYGSATVNSVPDFFDANKLTAITLVVDAGRRLTVRQLRFGNTVSADSTLRQEMRQEG 371
QY 381 ALASNQIKLSRARLMRTGFKHVTVDTR--PVPNSPDQVDVNVFVEQPGSGSTIAAGY 438
Db 372 TWYNSQLVELGKIRLDRGTGFFE--TVENRIDPINGSNDEVDVYKVKERNGTGSINFGIGY 429
QY 439 SOSGGVTFQFDVSONNFMCTGKHVNASFRRSETREYVSLGMTNPNYTVNGVVSQSLGYR 498
Db 430 GTESGISYQASVKQDNFLGTGAASVIACTKNDYGTSVNLGYTEPYFTKDG--SLGNGVF 487
QY 499 KTKYDNK---NISNVLDSDYSGSLSYGYPIDENORISFGL-----NADNTKLHGRFMGI 550
Db 488 FENYDNKSDTSNRYKRTYGSNLTGFPVNNNSYVGLGHTYKNKISNFALEYNRNLYI 547
QY 551 SNVKQLMADGKGIQVDNNGIPDPFKHDYTYNAILGNWYSSLDPRVFPQCGSHSVD--LT 608
Db 548 QSMK-----FKGNGIKTNDDFS-----FGWNYNSLRGYFTPKGVKASLGGGRVT 592
QY 609 VGFGDKTHOKVYVQGNIRPFE-----IKKSVLRGYAKLGYGN-NLPFFYENFVAGGYGS 660
```



Db 593 IPGSDNRYKLSADVOGFYPLDRDLHWVVSASAGYAN-GFGNKRPLFPYQTYTAGGIGS 651  
Qy 661 VRGYDQSSGLSPRSQAYLTARRGQQTTLG-----EVVGGNALATFGSELILPLPFKGDW 713  
Db 652 LRGFAYGSGIPNA---IYAEYNGSGTGTFFKISSDVLGGNAIATASAEELIVPTPFVSDK 708  
Qy 714 I-DQVRPVIFIEGGQVFDTT-GMDKQTDLTQFKDPQATAEONAKAANRPLLTQDKQLRY 771  
Db 709 SONTVRTSLFVDAASVWNTKWKSDKNGLSDVLK-----RLPDYGKSSRIRA.755  
Qy 772 SAGVGATWYTPIGPLSISYAKPLNKKQNDQTDVQFQIGSVF 813  
Db 756 STGVGFQWQSPIGPLVFSYAKPIKKYENDVDEQFQFSGGSF 797

## RESULT 3

US-08-433-522A-6  
; Sequence 6, Application US/08433522A  
; Patent No. 6013514  
; GENERAL INFORMATION:  
; APPLICANT: CHONG, Pele  
; APPLICANT: THOMAS, Wayne  
; APPLICANT: YANG, Yan Ping  
; APPLICANT: LOOSMORE, Sheena  
; APPLICANT: SIA, Dwo Yuan Charles  
; APPLICANT: KLEIN, Michel  
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN  
; NUMBER OF SEQUENCES: 55  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6TH Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/433,522A  
; FILING DATE: 12-SEP-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: STEWART, Michael I  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-434 MIS:jb  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 797 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein

US-08-433-522A-6

Query Match 25.08; Score 1052.5; DB 3; Length 797;  
Best Local Similarity 31.6%; Pred. No. 4.5e-76;  
Matches 260; Conservative 160; Mismatches 333; Indels 69; Gaps 20;

Qy 23 STHQAADFMANDITITGLQRVTIESQSLVPLFRIGQVVSQNLADGVKALYATGNFSDV 82  
Db 14 TTTVFAAPFVAKDIRVDQVGDLEQQIRASLPVRAGQRTDNDVANIVRSLFVSGRFDDV 73  
Qy 83 QVYHOGRI-IYQYTERPLAIEFEGNRLIPKEGLKAGNAGVQPLKQATVQVIE 141  
Db 74 KA-HOEGDVLVSVVAKSIISDIKGNVSIPTALKONLDANGFKVGDVIREKLNEFA 132

Qy 142 TELTNOYISQYNTETVTKOTMLDGNRVKQDMTFAEGKPARVVDVINIIGNQHSADADI 201  
Db 133 KSVKEHYASVGRNATVEPIVNTLPNRAEILIQINEDDKAKLASITFKGNESVSSSTLQ 192  
Qy 202 DVLAIKDNKINPLSKADRYTQEKLVTSLENRAKYLNGAFVRFPEIKDAKLINEDKNRIF 261  
Db 193 EOMELQDPSMWKL-WGNKFEQAQFEKDLQSDRYLLNNGYAKAQITKTQDVOLNDEKTKVN 251  
Qy 262 VEISLHGEQYRFQGTQFLGNLTYYOAELEALL-KFAEEGFSQAMLEQTTNNISTKFGD 320  
Db 252 VTIDVNEGLQDLRSARIIGNLGGMSAELEPLLSALHNDTFRSDIADVEMAIKALGE 311  
Qy 321 DGYTAAQIRPVTRINDESRITVDVEYIIDPVHPVYVRRINFTGNFKTQDEVLRMRQLEG 380  
Db 312 RYGSATVNSVPDEDDANKTLAITLVVDAGRLTVRQLRFEQNTVSADSTLRQENRQOEG 371  
Qy 381 ALASNQIKQLSRALMRTPGFFKHVTVDR--PVPNSPQVDVNVFVVEQSPSSSIIAGY 438  
Db 372 TWYNSQLVELGKIRLDRDTGFFE--TVENRIDPINGSNDEVDVYVKVKERTNGSINFGIGY 429  
Qy 439 SOSGGVTFQFDVSONNFMGTGKHVNASFSSRSETRVYSLGMTNPYFTVNGVVSQSLSGYVR 498  
Db 430 GTESGISYQASVKQDNFLCTGNAVSIAGTKNDYGTSVNLGYTEPYFTKDG--SLGGNVF 487  
Qy 499 KTKYDNK---NISNYVLDYSGSLSYGYPIDENQRISFCL-----NADNTKLHGGREFMI 550  
Db 488 FENYDNKSDTSSNYKRTYGTGNSVTLPFPVNNNSYVVLGHTYKNKISNFALEYNRNLYI 547  
Qy 551 SNVQLMADGGKIQVDNNGIPDFKHDTYTNAILGNWYSSLDPRVPFPTQGNHSHVD--LT 608  
Db 548 QSMK-----FKNGIKITNDFDS-----FCWNTNSLNRGYFPTKGKVASLGGRTV 592  
Qy 609 VFGDKTHQKVYQGNIVRPF-----IKKSVLRGYAKLGYGN-NLPYENFYAGGYGS 660  
Db 593 IPGSDNRYKLSADVOGFYPLDRDLHWVVSASAGYAN-GFGNKRPLFPYQTYTAGGIGS 651  
Qy 661 VRGYDQSSGLSPRSQAYLTARRGQQTTLG-----EVVGGNALATFGSELILPLPFKGDW 713  
Db 652 LRGFAYGSGIPNA---IYAEYNGSGTGTFFKISSDVLGGNAIATASAEELIVPTPFVSDK 708  
Qy 714 I-DQVRPVIFIEGGQVFDTT-GMDKQTDLTQFKDPQATAEONAKAANRPLLTQDKQLRY 771  
Db 709 SONTVRTSLFVDAASVWNTKWKSDKNGLSDVLK-----RLPDYGKSSRIRA 755  
Qy 772 SAGVGATWYTPIGPLSISYAKPLNKKQNDQTDVQFQIGSVF 813  
Db 756 STGVGFQWQSPIGPLVFSYAKPIKKYENDVDEQFQFSGGSF 797

## RESULT 4

US-09-135-166-2  
; Sequence 2, Application US/09135166  
; Patent No. 6083743  
; GENERAL INFORMATION:  
; APPLICANT: CHONG, Pele  
; APPLICANT: THOMAS, Wayne  
; APPLICANT: YANG, Yan Ping  
; APPLICANT: LOOSMORE, Sheena  
; APPLICANT: SIA, Dwo Yuan Charles  
; APPLICANT: KLEIN, Michel  
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN  
; NUMBER OF SEQUENCES: 55  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6TH Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

Db 74 KA-HQEGDVLVSVVAKSIIISDVKIKGNSVIPTEALKQNLNDANGFKVGDLIREKLNEFA 13

Db 593 IPGSDNKYYKLSADVQGFYPLDRDHLVWVSAKASAGYAN-GFGNKRLPFYQTYTAGGIGS 651



QY 661 VRGYDQSSLSGRQAYLTARRGQTTTG-----EVVGGNALATFGSELILPLPFKGDW 713  
Db 652 LRGFAYGSGIPNA---IYAEYNGSGTGTFFKKISSDVIGGNATATASAEILVPTFPFVSDK 708  
QY 714 I-DQVRPVIETEGGQVFDTT-GMDKQTDLTQFKDPQATAEQNAKAAANRPLLTQDKOLRY 771  
Db 709 SONTVRTSLFVDAASVWNTKWSKNGLESVDLK-----RLPDYKGSRRIRA 755  
QY 772 SAGVGATWYTPIGPLSISYAKPLNKKNDQDQTDVQFQIGSVF 813  
Db 756 STGVGFQWQSPIGPLVFSYAKPIKKYENDDVQEQFQSIGGSF 797

RESULT 7  
US-08-942-046-2  
; Sequence 2, Application US/08942046  
; Patent No. 6264954  
; GENERAL INFORMATION:  
; APPLICANT: CHONG, Pele  
; APPLICANT: THOMAS, Wayne  
; APPLICANT: YANG, Yan Ping  
; APPLICANT: LOOSMORE, Sheena  
; APPLICANT: SIA, Dwo Yuan Charles  
; APPLICANT: KLEIN, Michel  
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN  
; NUMBER OF SEQUENCES: 55  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6TH Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/942.046  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/433,522  
; FILING DATE: 12-SEP-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: STEWART, Michael I  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-732 MIS:jb  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 797 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-942-046-2

Query Match 25.0%; Score 1052.5; DB 4; Length 797;  
Best Local Similarity 31.6%; Pred. No. 4.5e-76;  
Matches 260; Conservative 160; Mismatches 333; Indels 69; Gaps 20;

QY 23 STHAQAAQFMANDITITGLQRTVIESLQSLVPLFRLCQVVSQNLADGVKALYATGNFSDV 82  
Db 14 TTTTFAAPFAVDKIRVDGVQGGLEQIIRASLPVRAGQRTDNDVANIVRSFVSGRFDDV 73  
QY 83 QYTHQEGRI-IYQVTERPLIAINEFGRNLIPKEGLQGLKKNAGLAVGQPLKQATVQMIE 141

Db 74 KA-HQEGDVLVSVVAKSIIISDVKIKGNSVIPTKALQNLDANGFKVGDVLIIRKLENEFA 132  
QY 142 TELTNOYISOGYYNTEITVAKOTMLDGNRKVKLDMTFAEGKPARVVDINIIGNHESDADLI 201  
Db 133 KSVKEHVASGRYNATVEPIVNTLPNNRAEILIOINEDDRAKASLTFKGNESVSSSTLQ 192  
QY 202 DVLAIAKDNKINPLSKADRYTQEKLVTSLENLRAKYLNAGVFVRFEIKDAKLINEDKNRIF 261  
Db 193 EQMELQDPSWWKL-WGNKFEQAQEKDLQSRDYLYLNGYAKAQITTKTDVGLNDEKTKVN 251  
QY 262 VEISLHGEQYRFQGTQFLNLTQAELEALLK-FRAEEGFSQAMLEQTTNNISTKFGD 320  
Db 252 VTIDVNEGLQDLRSARIIGNLGMASAEPLLSALHLNDTFRSDIADVENAKAKLIGE 311  
QY 321 DGYVYAOIRPVTRINDESRTVDVEYIDPVHPVVRVRRINFTGNFKTQDEVLRMRQLEG 380  
Db 312 RYGSATVNSVPDFFDANKTLAILVDVAGRRLIVRQLRFEGNTVSADSTLRQEMRQEG 371  
QY 381 ALASNQIQLSRARLMRTGFFKHVTVDTTR--PVPNSPDQVDVNFVVEPQPSSTIAAGY 438  
Db 372 TWNSQLVELGKIRLDRGTGFFE--TVENRIDPINGSNDEVDVYKVKERNGTSGINFGIGY 429  
QY 439 SOSGVTFQFDVSONNPMGTGKHVNASFRTSETRVYSLGTMNPXYETVNGYSQSLGYR 498  
Db 430 GTESGISYQASVKQDNFLGTGAASVIACTKNDYGTSMNLGTYNFKISFALEYNRNLYI 487  
QY 499 KTKYDNK---NISNVLDSDYSGSLSYGYPIDENORISFGL-----NADNTKLHGGRFMI 550  
Db 488 FENYDNKSDTSSNYKRTYGSNTVLTGPPVNNNSYVGLGHTYNTKISFALEYNRNLYI 547  
QY 551 SNVKQLMADGGKIQVNDNGIPDFKHDTYTYNAILGNWYSSLDPRVFPFQGMHSHVD--LT 608  
Db 548 QSMK-----FKNGIKTNDFDES-----FGWNTNSLNRGYFTPKGVKASLGGGRVT 592  
QY 609 VGFQDKTHQVYVYQGNLYRPF-----IKKSVLRGYAKLGYGN-NLPFYEYFAGYGS 660  
Db 593 IPGSDNKYIKLSADVOGFYPLDRDLHWVSASAKAGYAN-GFGNKKLFFYQTYTAGGIGS 651  
QY 661 VRGYDQSSLSGRQAYLTARRGQTTTG-----EVVGGNALATFGSELILPLPFKGDW 713  
Db 652 LRGFAYGSGIPNA---IYAEYNGSGTGTFFKKISSDVIGGNATATASAEILVPTFPFVSDK 708  
QY 714 I-DQVRPVIETEGGQVFDTT-GMDKQTDLTQFKDPQATAEQNAKAAANRPLLTQDKOLRY 771  
Db 709 SONTVRTSLFVDAASVWNTKWSKNGLESVDLK-----RLPDYKGSRRIRA 755  
QY 772 SAGVGATWYTPIGPLSISYAKPLNKKNDQDQTDVQFQIGSVF 813  
Db 756 STGVGFQWQSPIGPLVFSYAKPIKKYENDDVQEQFQSIGGSF 797

RESULT 8  
US-08-942-046-4  
; Sequence 4, Application US/08942046  
; Patent No. 6264954  
; GENERAL INFORMATION:  
; APPLICANT: CHONG, Pele  
; APPLICANT: THOMAS, Wayne  
; APPLICANT: YANG, Yan Ping  
; APPLICANT: LOOSMORE, Sheena  
; APPLICANT: SIA, Dwo Yuan Charles  
; APPLICANT: KLEIN, Michel  
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN  
; NUMBER OF SEQUENCES: 55  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6TH Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk





```
Db 652 LRGFAYGSGPNA---IYAEHNGTFFNKISSDVIGGNAITTAASAEILVPTFVSDKSQNT 708
Qy 717 VRPVIFIEGGVDFDT-GMDKQTDLTQFKDQPAQAEQNAKAAANRPLLTQDKQLRYSAGV 775
Db 709 VRTSLFVDAASVWNTKWSKNGLESKVLKD-----LPDYGKSSRIRASTGV 755
Qy 776 GATWYTPIGPLSISYAKPLNKKQNDQTDTVQFQIGSVF 813
Db 756 GFQWQSPIGPLVFSYAKPIKKYENDDDVEQFQFSGISGF 793

RESULT 11
US-09-135-166-10
; Sequence 10, Application US/09135166
; Patent No. 6083743
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/135,166
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 793 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-135-166-10

Query Match 24.98; Score 1047.5; DB 3; Length 793;
Best Local Similarity 31.78; Pred. No. 1.1e-75;
Matches 259; Conservative 160; Mismatches 334; Indels 65; Gaps 21;

Qy 23 STHQAADFMANDITITGLQRTTSTESLVPFLRGVQVSVENQLADGVKALYATGNFSDV 82
Db 14 TTTFAAPFVPKRDVQGDLEQIRASLFRAGQRTDNDVNIYRSLFVSGRFDDV 73
Qy 83 QVYHQEGRI-IVQVTERPLAIENFEGNRLIPKPGQLGSLKAGLAVGQPLKQATVQMI 141
Db 74 KA-HQEGDVLVVSVVAKSIISDVKIGNSVIPTEALKQNLQDANGFKVGDVLIREKLNFEA 132
```

```
Qy 142 TELTNOYISQYNTETITVKTMLDGNRVKLDMTFAEGKPARVVDINIIGNQHFSADALI 201
Db 133 KSVKEHYASVGRNATVEPIVNTLPNNRAEILIOINEDDKAKLASITFKGNSVSSSTLQ 192
Qy 202 DVLATKDKINKPLSKADRYTQEKLVTSLENRAKYLNAGFVRFEEKADKLINEDKNRIF 261
Db 193 EOMELQPSWMLK-WGNKFEQAQFEKDLQAIIRDYLLNNGYAKAQITKTQVQLNDEKTKYN 251
Qy 262 VEISLHEGEQVRFQGTQFLGNLTYYTQAELEALLK-FKAEEGFSQAMLEQTTNNISFKGD 320
Db 252 VTIDNGLQYDLRSARIIGNLGGMSAELEPLLSALHLNLTFFRSDIADVENAIKAKLGE 311
Qy 321 DGYVYAIQRPVTRINDESRITVDVEYIIDPVHVVYVRRINFTGNFTQDQDLVLRMRQLSG 380
Db 312.RGYGNTTVNSVPDFDDANKTLAITFVVDAGRRLTVRQLRFEGNTVSADSTLQENRQBSG 371
Qy 381 ALASNQKIQLSRARLMRTGFFKHVTVDIR--PVPNSPOVDVNVFVVEQPSGSSSTIAAGY 438
Db 372 TWYNSQLVELGKIRLDRTGFFE--TVENRIDPINGSNDEVDVYVVKVKTERTGSINFGIGY 429
Qy 439 SQSGGVTFQFDVSONNFMGTGKHVNASFSRSETREVYISLGMTNPYFTVNGVSQSLSGYYR 498
Db 430 GTESGISYQTSIKQDNFLCTGAANSIAGTKNDYGTSVNLGYTEPYFTKDV--SLGNIF 487
Qy 499 KTYDNK---NISNYLDSYGSLSYGPIDENQRISFQL-----NADNTKLHGGRFMGI 550
Db 488 PENYDNSKSDTSSNYKRTTYGNSVTLGFPVNNNSYVVLGHTYKNISNFALEYNNLYI 547
Qy 551 SNVKOLMADGGKIQVDNNGIPDFKHDYTYNAILGNWYSSSLDRPVPPTQGMHSVD--LT 608
Db 548 QSMK-----FKNGIKTNDFFS-----FGWNTNSLNRGYFTPKGVKASLGGRYT 592
Qy 609 VGFGDKTHQKV--YOG-----NIYRPTFKSVLRGYAKLGYGN-NLPPYENFYAGGYGS 660
Db 593 IPGSDNKYKLSADVQGFYPLDRDRHWVWSAKASAGYAN-GGKNRLPPYQYTTAGGTGS 651
Qy 661 VRGYDQSSGLSPRSQAYLTARRQQT---TLGEVVGGNALATFGSELILPLPKGDMI-DO 716
Db 652 LRGFAYGSGPNA---IYAEHNGTFFNKISSDVIGGNAITTAASAEILVPTFVSDKSQNT 708
Qy 717 VRPVIFIEGGVDFDT-GMDKQTDLTQFKDQPAQAEQNAKAAANRPLLTQDKQLRYSAGV 775
Db 709 VRTSLFVDAASVWNTKWSKNGLESKVLKD-----LPDYGKSSRIRASTGV 755
Qy 776 GATWYTPIGPLSISYAKPLNKKQNDQTDTVQFQIGSVF 813
Db 756 GFQWQSPIGPLVFSYAKPIKKYENDDDVEQFQFSGISGF 793

RESULT 12
US-08-942-046-10
; Sequence 10, Application US/08942046
; Patent No. 6264554
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

		SOFTWARE: PatentIn Release #1.0, Version #1.25	
		CURRENT APPLICATION DATA:	
		APPLICATION NUMBER: US/08/942,046	
		FILING DATE:	
		CLASSIFICATION: 435	
		PRIOR APPLICATION DATA:	
		APPLICATION NUMBER: 08/433,522	
		FILING DATE: 12-SEP-1995	
		CLASSIFICATION: 435	
		ATTORNEY/AGENT INFORMATION:	
		NAME: STEWART, Michael I	
		REGISTRATION NUMBER: 24,973	
		REFERENCE/DOCKET NUMBER: 1038-732 MIS:jb	
		TELEPHONE: (416) 595-1155	
		TELEFAX: (416) 595-1163	
		INFORMATION FOR SEQ ID NO: 10:	
		SEQUENCE CHARACTERISTICS:	
		LENGTH: 793 amino acids	
		TYPE: amino acid	
		TOPOLOGY: linear	
		MOLECULE TYPE: protein	
		US-08-942-046-10	
		Query Match 24.9%; Score 1047.5; DB 4; Length 793;	
		Best Local Similarity 31.7%; Pred. No. 1.1e-75;	
		Matches 259; Conservative 160; Mismatches 334; Indels 65; Gaps 21;	
QY	23	STHAQAADFMAADITITGLQVTTIESLQSVLPFRIGLVVSENOLADGVKALYATGNFSDV	82
DB	14	TTTFAAPFPKIRVDVGVQDLEQIRASLPVRAGQRTDNDVANIIVRSFLVSGRFDV	73
QY	83	QYVHOGRI-IYQVTERPLIAEINFEGRNLPKQGLQKLNAGLAVQPLKQATVOMIE	141
DB	74	KA-HOEGDVLVSVVAKSIISDVKIKGNSVIPTKALQKNDANGPKVGDVILIREKLNFEA	132
QY	142	TELTNQYISQGYNTYNTVKTMLDGNRKLDMTFAEKGPARVVDINIGNHFSADLI	201
DB	133	KSVKEHYASVGRYNATVEPIVNTLPNNRAEILIQINEDDKAKLASITFKGNEVSSTLQ	192
QY	202	DVLAIKDNKINPLSKADRYTOEKLVTSLNRAKYLNAGFVRFPEIKDAKLINEDKNRIF	261
DB	193	EQMELQPSWKKL-WGNKFEQAQFEKDLQAIRDYLLNNGYAKAQITKTDVOLDNKTKN	251
QY	262	VEISLHEGBOYRFGOTQFLGNLTYYQAELEALLK-FKAEEGFSQAMLEQTNNISTKFGD	320
DB	252	VTDVNEGQYDLRSARIIIGNLGGMSAEPLLSALHLNDFRSDIADVENAIKALGE	311
QY	321	DGYTAAQIRPVTRINDESRVDVEYIIDVHPVYVRRINFTGNFKTQDEVLRMRQLEG	380
DB	312	RCYGNNTVNSVPDEDDANKTLAITFVVDAGRRLTVROLRFEGNTVSADSTLRQEMRQBG	371
QY	381	ALASQKIQLSARLMRTGFFKHVTVDR--VPVNSPDVDYVNVVEEQPSGSSTIAAGY	438
DB	372	TWNSQLVELGRIDRTGFFE--TVENRIDPINGSNDEVDVYKVKERTGNSIFNGIGY	429
QY	439	SOSGGVTFQFQYSONFMGTGKHVNASFSRSTREYISLGMTNPNFTVNGVSQSLSGYR	498
DB	430	GTEGSIYQTSYTKQDNFLGTGAIVSTAGTKNDYGSVNLGYTEPYTKDGV--SLGGNIF	487
QY	499	KTKYDNK---NFSNYLDSYSGSLSYGYPIDENQRIISFGL-----NADNTKLHGGRFMI	550
DB	488	FENYDMSKSDTSNYKRTTYGNSVTLGFPVNNNSYVGLGHTYNNKISNFALEYNNRLYI	547
QY	551	SNVKQLMADGGKIQVDNNGIPDKHDTYTNALGHNYSLLDRPVPPTQCMHSVD--LT	608
DB	548	QSMK-----FKNGIKTNDFFS;-----FGWNSLNKRGYPTTKGVKASLGRVT	592
QY	609	VGFQDTHOKV--YQG-----NIYRPFKIKSVLRGAKLGYGN-NLPFYENFYAGGYS	660
DB	593	IPGSDNKKYKLSADVQGFPLDRDRHWVWSAKASAGYAN-GFGNKRLLPFQVTVTAGIGS	651











GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 09:51:21 ; Search time 69.95 Seconds  
(without alignments)  
1116.807 Million cell updates/sec

Title: US-09-701-711-2  
Perfect score: 4202  
Sequence: 1 MRNSYFKGFQVSAMTMAVMM.....LNKKQNDQDTVQFQIGSVF 813  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues  
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_71.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Length DB ID	Description
1	1451.5	34.5	797 2 H83190	probable outer mem
2	1262	30.0	795 2 AC0129	probable surface a
3	1218.5	29.0	803 2 AB0530	outer membrane pro
4	1210	28.8	810 2 A64742	hypothetical prote
5	1210	28.8	810 2 C90651	hypothetical prote
6	1210	28.8	810 2 C85502	hypothetical prote
7	1203	28.6	803 2 B82099	surface antigen VC
8	1172.5	27.9	784 2 E82731	outer membrane ant
9	1115	26.5	797 2 C81228	outer membrane pro
10	1111	26.4	797 2 D82000	outer membrane pro
11	1053.5	25.1	808 2 F64102	protective surface
12	1052.5	25.0	797 2 JC4078	protective surface
13	754	17.9	768 2 D17126	outer membrane pro
14	734	17.5	774 2 D97527	omp1 protein precu
15	734	17.5	774 2 AE2746	group 1 outer memb
16	727.5	17.3	769 2 F87486	outer membrane pro
17	697	16.6	768 2 B97725	outer membrane pro
18	680.5	16.2	781 2 AH3355	outer membrane pro
19	641	15.3	617 2 H84957	hypothetical prote
20	568.5	13.5	778 2 C70412	outer membrane pro
21	515	12.3	739 2 A81430	outer membrane pro
22	434	10.3	916 2 G64601	protective surface
23	429	10.2	906 2 F71910	probable outer mem
24	397.5	9.5	792 2 H81693	outer membrane pro
25	396.5	9.4	792 2 B71539	probable omp85 ana
26	395	9.4	790 2 D86528	omp85 analog [impo
27	395	9.4	790 2 D72094	omp85 analog - Chl
28	336.5	8.5	853 2 A71339	probable outer mem
29	325	7.7	821 2 B70199	outer membrane pro

ALIGNMENTS

RESULT 1

H83190

Probable outer membrane protein PA3648 [imported] - Pseudomonas aeruginosa (strain P  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: H83190  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.;  
Lory, S.; Olson, M.V.  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa  
Nature 406, 959-964, 2000  
A:Reference number: AB2950; MUID:20437337  
A:Accession: H83190  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-797 <STO>  
A:Cross-references: GB:AB004784; GB:AE004091; NID:g9949799; PIDN:AAG07036.1; GSPDB:G  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA3648  
C:Superfamily: protective surface antigen D-15

Query Match				
Best Local Similarity	34.5%	Score 1451.5;	DB 2;	Length 797;
Matches 308;	Conservative 163;	Mismatches 297;	Indels 67;	Gaps 13;
Qy	7	KGQVSAMTMVMMYMSHAQAADFMANDITITGLQRTVIESLSQSVLPFRLGQVSVENOL	66	
Db	2	KRFLPAL-LSALMAIEVHAES--FTVSDIRVNGLRVSAGSVFAALPLNVGTETDDQAL	58	
Qy	67	ADGVKALVATGNFSQVYHQEGRIYQVTERPLTAENFEGNRLIPKGLGEGLNAGL	126	
Db	59	VOATRSFKTGFQDIQLRGDNVLVTVVVERPSISSIEIEGNKAISKEDLLKLGKQSL	118	
Qy	127	AVGQPLKQATVQMIETLTNOYISQYNTETITVQTMLDGNRKLDMTFAGKPARVVD	186	
Db	119	AGEGFQKATLEGVNEQRQVVAQGRYSAEINAEVQPPNRNRVALKININGTVNAISH	178	
Qy	187	INIQNHFSADLDIVLAIK-DNKINPLSKADRYTQEKLVTSLENLRKRYLNAGVRFEE	245	
Db	179	INVGVNTVFSBEDTDLFELKTTNLSFFKNDDKYAREKLSGDLERLSRYLDRGVINMD	238	
Qy	246	TKDAKLNTNEDKNTFVEISLHEGEQYRFGQTFQGNLTYYTQAELEALKFAEGBFSA	305	
Db	239	IATQVSITPDKKHVYITVINEGEKTYTRDYKLTGDLKVPBEEVKRLLLVKGQGVFSRK	298	
Qy	306	MLEQTTNNISTKFGDDGYVYIAQIRPVTRINDESRVDVEYIDPVHPVVRVRRINFTGNFK	365	
Db	299	VMTTSDLTITRLNGEGTTFANVGPVPEAHDDDKTVSVTFVVDGPKRAVNRINFRGNTK	358	
Qy	366	TQDEVLRREMROLEGALASNQIQLSRARLMRTGFFKHVTVDTRPVPSPDQDVNFVVE	425	

Db 359 TEDEVLRREMRQEGGWASLYLIDQSKARLERLGKFEVNEVPANPGTDQDVVNSVE 418  
QY 426 EQPSGSGSTIAAGYSQSGVTFQFDVSONNPMGTCKKHYNASFSRSETRVYSLGNTDNYFT 485  
Db 419 EQPSGSGSTIAAGYSQSGVTFQFDVSONNPMGTCKKHYNASFSRSETRVYSLGNTDNYFT 478  
QY 486 VNGVQSGLSGYRKTVD--NKNISNVYLDVSDVSGSLSGYGYPIDENQRIISGLNADNKLH 543  
Db 479 VDGVSGLGNAYFKRTDYDELVDVDSVNSVSLGMSIGYPISTSLTTLGLSVQRDQID 538  
QY 544 GGRFMGTSNVKKLMADGGKIQVDNNGIPDFKHDTYTNAILGNWYSILDRPVPFTQGMH 603  
Db 539 TGRIT-VDEIYDFLDKES-----DN-----PTNFKASIGWSESTLNGVLATRGHSQ 584  
QY 604 SVDL--TVGFGDKTHQKVVYOGNIYRPFKKSVLRGYAKLGYGN-----NLPEYENFYA 655  
Db 585 SLTLETLPGSGLSYKIDYRGVAFPLTDNYTRFHTELGYGDSYSTERLPYENFYA 644  
QY 656 GGYGSRVGYDSSGLGPRSOAVLTARRGQQTTL-----GEVWGGNALAT 698  
Db 645 GGFNSVRGFKDSTLGRSTPVS-ARNDGTPKKNQGPSKGRYTDPODPAFGNLTIT 703  
QY 699 FGSLLILPLPFKGDWIDQVREVFIEGGQVEDTTGMDKQITDLTQFKDPQATASQNAKAA 758  
Db 704 GGAELLPLPFVKDQ-RQLRTVLFWDVGSSTEDTDCPTKTTTNCDBGIK-----749  
QY 759 NRPLLTQDKQLRYSGAGVATWTPIGPLSISYAKPLNKKONDQDTVQFOIGSVF 813  
Db 750 -----TDNLASSVGVGLTWITALGPLSLATPIKPPDNAETQVQFSLGQTF 797

RESULT 2  
AC0129  
probable surface antigen YP01052 [imported] - Yersinia pestis (strain C092)  
C:Species: Yersinia pestis  
C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Nov-2001  
C:Accession: AC0129  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;  
geno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
li, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrrell,  
Nature 413, 523-527, 2001  
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AC0129  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-795 <KUR>  
A:Cross-references: GB:AL590842; PIDN:CAC89894.1; PID:g159791119; GSPDB:GN00175  
C:Gene: YP01052  
C:Superfamily: protective surface antigen D-15

Query Match 30.0%; Score 1262; DB 2; Length 795;  
Best Local Similarity 35.7%; Pred. No. 3.8e-65;  
Matches 292; Conservative 152; Mismatches 329; Indels 44; Gaps 14;  
QY 14 MTMAVMMVMS-THAQAD-FMANDITITGLQRTVIESLSQVLPFRGQVYSENOLADGVKA 72  
Db 6 LIASLLSFSATVYAGDFVNDIHFEGLRVAVGAALLNMPVRVGTVDSDDDIGKTIRA 65  
QY 73 LYATGNFSDVQVHQEGRIYQVTERPLIAEINFEGRNRLIPKEGLQEGLNAGLAVGQPL 132  
Db 66 LEATGNFEDVRLDGNLTILVQVKRTIASITFSGNKAVKEDMDIKONLEASGVRVGEAL 125  
QY 133 KQATVOMIETELNOYISQGYNTETITVKTMDGNRKLDMPTFAEGKPARVVDINIIGN 192  
Db 126 DRTTISNIEKLEDEGKSGVKSASVAVVTPLRNRVDLKVFTTEGVSAKIQIINLVGN 185  
QY 193 QHPSADLIDVLAIFD-----NKNIPLSKADR-YTOEKLVTSLENRAKYLNAGFVRFEI 246  
Db 186 HSTTDELIISRFQDRVDPWNNV-----GDRKYQKOKLAGDLETLRSFLDRGYARFNI 240

QY 247 KDAKLINEDKNRIFVEISLHEGEQYRFQOTFLGNLTYTQAELEALLKFAEGBFSQAM 306  
Db 241 DSTQVSLTPDKKGIYVITINTEGPOFKLSNVISGNLAGHQSEAKLTKEPGLFNGSK 300  
QY 307 LEQTTNISTKFGDDGYIAQIRPVTRINDESRIVDVEYKIDPVHVPVYVIRINFTGNEKT 366  
Db 301 VTRMEDDIKMLGRYGYAYPRVVTQPEINDDDKTKLHINVDAGNRFRVRIHREGNDTS 360  
QY 367 QDEVLRREMRQEGGWASLYLIDQSKARLERLGKFEVNEVPANPGTDQDVVNSVE 426  
Db 361 KDSVLRREMRQEGGWASLYLIDQSKARLERLGKFEVNEVPANPGTDQDVVNSVE 420  
QY 427 QPSGSGSTIAAGYSQSGVTFQFDVSONNPMGTCKKHYNASFSRSETRVYSLGNTDNYFT 486  
Db 421 RNTGSLNFGIGYGTESGVSFQVQVQDNWLTGNTVINGTKNDYQTYAEFTLMDPYFTV 480  
QY 487 NGVSQSLSGYRKTVDNKNISNVYLDVSDVSGSLSGYGYPIDENQRIISGLNADNKLH 546  
Db 481 DGVSLGGRIFTNDFKADNADSLGTYNSYAGDGLGFFPINENNSLRVGVYVHNDL-SDM 539  
QY 547 FMGISNVKQLMADGGKIQVDNNGIPDFKHDTYTNAILGNWYSILDRPVPFTQGMHSDV 606  
Db 540 LPQVAMWRYLESVGERPGYD--GREGFTTDDFTLN--LGWTYNNLDRGFFFTSGVKSNN 595  
QY 607 --LTVGFGDKTHQKVVYOGNIYRPF--IKKSVLRGYAKLGYGN-----NLPEYENFYAGG 657  
Db 596 TKITVPGSDNEFYKVTFTDSAYQPLNEDRSWVLLGRGLGYGDTGSKEMPEYENFYAGG 655  
QY 658 YGSRVGYDSSGLGPRSOAVLTARRGQQTTLGVEVGGNALATFGESELIPLPFKGD-WIDQ 716  
Db 656 SSTVRGFRSNNILGPKA-AYANGGATVNTSDAVGNAVAVASIELITPTPFISEKYSNS 714  
QY 717 VRPVIFIEGGQVDFDTGMDKQITDLTQFKDPQATASQNAKARNRPLLTQDKQLRYSGAVG 776  
Db 715 VRTSIFDSGTWMDTNW-----ENTAK--TRAAGIPDYGKASNIRVSAGVA 758  
QY 777 ATWYTPIGPLSISYAKPLNKKONDQDTVQFOIGSVF 813  
Db 759 LQWMSPLGPLVFSYAKPVKDYEGDKSEQQFQFNIGTW 795

RESULT 3  
AB0530  
outer membrane protein precursor yaer [imported] - Salmonella enterica subsp. enteric  
C:Species: Salmonella enterica subsp. enterica serovar Typhi  
A:Note: This species has also been called Salmonella typhi  
C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AB0530  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church  
th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Fair  
, S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,  
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se  
A:Reference number: AB0502; PMID:11677608  
A:Accession: AB0530  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-803 <PAR>  
A:Cross-references: GB:AL513382; PIDN:CAD08682.1; PID:g16501505; GSPDB:GN00176  
C:Genetics:  
A:Gene: yaer  
C:Superfamily: protective surface antigen D-15

Query Match 29.0%; Score 1218.5; DB 2; Length 803;  
Best Local Similarity 33.9%; Pred. No. 1.2e-62;  
Matches 285; Conservative 150; Mismatches 327; Indels 79; Gaps 15;  
QY 14 MTMAVMMVMS-THAQAD-FMANDITITGLQRTVIESLSQVLPFRGQVYSENOLADGVKA 67  
Db 1 MAMKLLIASLLPSSATVYAGDFVNDIHFEGLRVAVGAALLNMPVRVGTVDSDDDIGKTIRA 60

[illegible]

sequence not shown; translation not shown

A: Cross-references: GB:AE000127; GB:U00096; NID:gl786370; PIDN:AAC73288.1; PID:gl786370  
 A: Experimental source: strain K-12, substrain MG1655  
 C: Superfamily: protective surface antigen D-15

Query Match	28.8%;	Score 1210;	DB 2;	Length 810;
Best Local Similarity	34.0%;	Pred. No. 3.9e-62;		
Matches 287;	Conservative 149;	Mismatches 330;	Indels 78;	Gaps 15;

Qy	14	MTMAVMYMS-----THAQADPMPANDITITIGLQRYTIESLQSVLPFRLLGQVVSENOLA	67
Db	1	MAKKKLLTASLLFSATVYAGEGVVWKIHFEGLGRVAVGAALLSMPVRTGDTVNDSDIS	60
Qy	68	DGVKALYATGNFSDVQVYHQBGRILYQVTERPLIAEINFEGRNLLIPKEGLQEBGLKNAGLA	127
Db	61	NITRALFATGFNFEDVRVLRDGDTLLVQKERTIASITFSGNKSVKDDMLKQNLASGVR	120
Qy	128	VQQLKQATQVMIETELFNQYISQYYNTETITVQOTMLDGNRVKLDMTFAEGKPARVVDI	187
Db	121	VGESLDRTTIADIEKGLEDFYSGVKSASVAVVTPLPNRNVLDKLIVQEGVSAEIQOI	180
Qy	198	NIIQNOFSDADLLDVLAIKD-----NKIPLSKADR- YTOEKLIVTSLEMLRAKYINAGF	241
Db	181	NIVNGHAFTTDELISHQLFQEDVPMWNV-----GDRKYQKLAGDLETLSRYSYLDRGY	235
Qy	242	VRFEITKAALMINDEKKNRIFVEISLHGEQYRFQOTQFLGNLTNYTQAELEALLKFAEAG	301
Db	236	ARFNIDSTQVSLTPDKGIYVYTNITGSDQYKLSGVEVSNLACHSAEIEQLTKIEPGEI	295
Qy	302	FSQMLEQTTNNISTKFGDGGYYVQAI RPVTRINDESRITVDVEYYIDPVHPVYVRRINF	361
Db	296	YNGTKVTKMEDDIIKLLGRGYAYPRVQSMPEINDADKTVKLRVNVADAGNRFYVRIRP	355
Qy	362	GNETQDEVLRREMRQLEGALASNOKIQLSARILMRTGFEKHVTVDTRPVPNSPODVYN	421
Db	356	GNDTSKAYLRREMRQEWGLGSDLVQDQKERLNRUGFEETVDTQRPVSGPODVVV	415
Qy	422	FVBEQPGSGSTIAAGYSQSGVTFFQFDVSNQFMGTGKHVNASFSETREYISLGMTN	481
Db	416	YKVERNTGPNFGIGYGTESGSFQAGVQODNWLGTGYAVGNGTKNDYQTYAELSVTN	475
Qy	482	PIYFVNGVSQSLSGYRKYKVDNKNISNYVLDSSYGSLSYGYPIDENQRISEGLNADNTK	541
Db	476	PIYFVNGVSLGRIFYNDFOADDLSUDYTKSYGTDVTLGTFPINYSNLRAGLGYVHNS	535
Qy	542	LHGGRFMGISNVKJLMA-----DGGKIQVDNNGIPDFKHDTYTYNAILGWNYSSLDREV	595
Db	536	L-----SNMQPOVAMRWLYLSMGEHPSTSDQNSFKTDDTFEN- YGWTYNKLDRCY	585
Qy	596	PPTOG--MSHSVDLTVGFQDKTHQVYVQGNITRPF--IKKSVLRYGAKLGYGNL----	647
Db	586	FPTDGSRVNLTGKVTIPGSONEYYKYTLDPATVPIPDDDKHVVLTGRTWGYDGLGGKE	645
Qy	648	-PFYENFYAGYGSVRGYDOSSLGPRS-----QA-----YLTARRGOQTLL-----GBV	690
Db	646	MPFYENFYAGSSTVRGFSQNTIGPRAVYFPHQASNYDPDYECATQDGAOKDLCKSDDA	705
Qy	691	VGGNALATFGSELILPLPFGKD--WIDQVRVPIPIEGGVQVDTFGMDKQIDITLQFKDQPA	749
Db	706	VGGNAMAVASLEFITPTPLISDRYANSVRTSFFDMGTWMDTNWDSQYSGYPDYSDP----	763
Qy	750	TAEQNAKAANRPLLTKDQILRYSAGVGATWYTPIGPLSLSYAKPLNKKQNDQTDIVQOPI	809
Db	764	-----SNTMSAGIALQMSPLGLPLVFYSYQPFKKYDGDRAEQFQFNI	806
Qy	810	GSVF	813
Db	807	GKTV	810

## RESULT 5

C90651  
hypothetical protein ECs0179 [imported] - *Escherichia coli* (strain O157 H7, substrain

cholia coli  
2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
Accession: C90651  
A:Species: *Escherichia coli*  
A:Strain: T.; Makino, K.; Onishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Sasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic reference strain O157:H7  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: C90651  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-810 <STO>  
A:Cross-references: GB:BA000007; PIDN:BA33602.1; PID:g13359635; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrate RIMD 0509952  
C:Genetics:  
A:Gene: ECs0179  
C:Superfamily: protective surface antigen D-15

Query Match 28.8%; Score 1210; DB 2; Length 810;  
Best Local Similarity 34.0%; Pred. No. 3.9e-62;  
Matches 287; Conservative 149; Mismatches 330; Indels 78; Gaps 15;

QY 14 MTAVMMVMS-----THAQAADFMANIDITITGLQRTVIESLSVLPFRGQVVSQNL 67  
DB 1 MAMKLLIASLSSATVYGAEGFVVKDIHFEGQLQAVAGAAALSMFVRTGDTVND 60  
QY 68 DGKALYATGNFSDVQVYHOGRIIYOVTERPLIAEFNFGNRLIPKLEGLOGLKNAGLA 127  
DB 61 NTRALFATGNFEDVRLRGDTLLVQKERTIASITFSGNKSVKDDMLKNLEASGR 120  
QY 128 VGQPLQATQVMIETELTNQISQYNTETITVKTMLDGNRVKLDMTFAEGKPARVVDI 187  
DB 121 VGESLDRITADIIEKLEDFYYSVGKYSASVAVVTPLRNRVLDKLVFQEGVSAEIQI 180  
QY 188 NITGNQHFSDADLIDVLAIKD-----NKINPLSKADR-YTOEKLVTSLNLRKYLNAGF 241  
DB 181 NIVGNHAFTHDELISHFQRLDERVPMWV-----GDRKYQKQKLAGDLETLSYLD 235  
QY 242 VRPEIKAKLINENKKNRIFVEISLHGEQYRFQGTQFLGNLTQAELEALLKFAE 301  
DB 236 ARFNIDTQVSLTPDKKGIYVTVNITGQYKLSGVEVSNLASHSAEIQLTKEP 295  
QY 302 FSQAMLEQTTNNITKFGDDGYIAQIRPVTRINDESRTVDVEYIDVHPVYVRRINFT 361  
DB 296 YNGTKVTKMEDDIKLLGRYGYAYPRVQSMPEINDADTKLVNRVNDAGNRFYVRKIR 355  
QY 362 GNFTQDEVLRREMRQLEGALASQKIQLSARLMRTGFFKHVTVDTTPVNSPDQVDV 421  
DB 356 GNDTSKDAVLRREMRQLEGALASQKIQLSARLMRTGFFKHVTVDTTPVNSPDQVDV 415  
QY 422 FVBEQPSGSGSTIAAGYSQSGVTFQFQVSONNPMGTGKHVNASFSSRSETREVYSLG 481  
DB 416 YKVERNTGSENFNGIGYTESGVQFQAGVQDQWNLGTGYAVGNGTKNDYQTYAELSV 475  
QY 482 PYFTVNGVQSLSGYKRTKYNKNISNYLDSYSGSLSYGYPIDENQRTISFGLNADNTK 541  
DB 476 PYFTVNGVSLGRFLFYNDFOADDADLSYTNKSYGTDVTLGFPINEXNSLRAGLGYV 535  
QY 542 LHGRFPWGISNVKOLMA-----DGGKIQVDNNGIPDPKHDTYTNALLGNYSLSR 595  
DB 536 L-----SNNQPOVAMRWLYLSMGEHPSTSDQNSFKTDFTFN--YGMTYKLRG 585  
QY 596 FPTQG--MSHSVDLTGFGDKTHQKVYVQGNIRPF--IKKSLRGYAKLGYGNL----- 647  
DB 586 FPDGSRVNLTKGTIPGSDNEFYKVTLDATYVPIDDDKHWVLGRTRWYGGDLG 645  
QY 648 -PFYENYAGGYGVRGDSLSGRS-----QA-----YLTARRGQQTTL---GEV 690  
DB 646 MPFYENYAGGSSTVRGFSQNTIGFKAIVFPHQASNDPDYDYECATQDGAACKLCKSD 705  
QY 691 VGNALATFGSELILPLPFKGD-WIDQVRPVIFTEGGQVDTTGMCKQTIDLTKQKPOA 749

Db 706 VGNAMAVASLEFPTTPTPFISDKYANSVRTSFFWDMGTVMNDWSSQYSGYPYSDP-- 763  
QY 750 TAEQNAKAANRPLLTQDKQLRYSAGVGATWYTPIGPLSISYAKPLNKKQNDQTDVQFI 809  
Db 764 -----SNIRMSAGIALQWMSPLGLVFSYAQPFKYDKGDKABQFQFNI 806  
QY 810 GSVF 813  
Db 807 GKTW 810  
RESULT 6  
C85502  
hypothetical protein yaeF [imported] - *Escherichia coli* (strain O157:H7, substrate ED  
C:Species: *Escherichia coli*  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C:Accession: C85502  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May  
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: C85502  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-810 <STO>  
A:Cross-references: GB:AE005174; NID:g125112902; PIDN:AAG54479.1; GSPDB:GN00145; UWGP:  
A:Experimental source: strain O157:H7, substrate EDL933  
C:Genetics:  
A:Gene: yaeF  
C:Superfamily: protective surface antigen D-15

Query Match 28.8%; Score 1210; DB 2; Length 810;  
Best Local Similarity 34.0%; Pred. No. 3.9e-62;  
Matches 287; Conservative 149; Mismatches 330; Indels 78; Gaps 15;

QY 14 MTAVMMVMS-----THAQAADFMANIDITITGLQRTVIESLSVLPFRGQVVSQNL 67  
DB 1 MAMKLLIASLSSATVYGAEGFVVKDIHFEGQLQAVAGAAALSMFVRTGDTVND 60  
QY 68 DGKALYATGNFSDVQVYHOGRIIYOVTERPLIAEFNFGNRLIPKLEGLOGLKNAGLA 127  
DB 61 NTRALFATGNFEDVRLRGDTLLVQKERTIASITFSGNKSVKDDMLKNLEASGR 120  
QY 128 VGQPLQATQVMIETELTNQISQYNTETITVKTMLDGNRVKLDMTFAEGKPARVVDI 187  
DB 121 VGESLDRITADIIEKLEDFYYSVGKYSASVAVVTPLRNRVLDKLVFQEGVSAEIQI 180  
QY 188 NITGNQHFSDADLIDVLAIKD-----NKINPLSKADR-YTOEKLVTSLNLRKYLNAGF 241  
DB 181 NIVGNHAFTHDELISHFQRLDERVPMWV-----GDRKYQKQKLAGDLETLSYLD 235  
QY 242 VRPEIKAKLINENKKNRIFVEISLHGEQYRFQGTQFLGNLTQAELEALLKFAE 301  
DB 236 ARFNIDTQVSLTPDKKGIYVTVNITGQYKLSGVEVSNLASHSAEIQLTKEP 295  
QY 302 FSQAMLEQTTNNITKFGDDGYIAQIRPVTRINDESRTVDVEYIDVHPVYVRRINFT 361  
DB 296 YNGTKVTKMEDDIKLLGRYGYAYPRVQSMPEINDADTKLVNRVNDAGNRFYVRKIR 355  
QY 362 GNFTQDEVLRREMRQLEGALASQKIQLSARLMRTGFFKHVTVDTTPVNSPDQVDV 421  
DB 356 GNDTSKDAVLRREMRQLEGALASQKIQLSARLMRTGFFKHVTVDTTPVNSPDQVDV 415  
QY 422 FVBEQPSGSGSTIAAGYSQSGVTFQFQVSONNPMGTGKHVNASFSSRSETREVYSLG 481  
DB 416 YKVERNTGSENFNGIGYTESGVQFQAGVQDQWNLGTGYAVGNGTKNDYQTYAELSV 475  
QY 482 PYFTVNGVQSLSGYKRTKYNKNISNYLDSYSGSLSYGYPIDENQRTISFGLNADNTK 541  
DB 476 PYFTVNGVSLGRFLFYNDFOADDADLSYTNKSYGTDVTLGFPINEXNSLRAGLGYV 535



```
Qy 542 LHGGFPMGISNVKQLMA-----DGGKIQVDNNGIPDPKHDYTTYNAILGNWYSSLDPRV 595
Db 536 L-----SNMQPQVAMRWLYLSMGEPSTSDQNSFKTDFTFN--YGTYNKLDRGY 585
Qy 596 FPTQO--MSHSVDLTNGFGDKTHOKVYVOGNIYRPE--IKKSVLRGYAKLGYGNL----- 647
Db 596 FPDGSRVNLTKVTIPGSDNEYKYVTLDTATYVPIDDDHKWVLRGTRWYGGDLGCKE 645
Qy 648 -PFYENFYAGGYSGVRGQDSSLGPRS-----QA-----YLTARRGQOQTTL---GEV 690
Db 646 MPFYENFYAGGSSTVRGFSQNTIGPKAVYFPHQASNYDPDYDECATQDGAKDLCSDDA 705
Qy 691 VGGNATATFGSELILPLPFFKGD-WIDQVRPVIFIEGGQVFTTGMKDKTIDLTQFKDPQA 749
Db 706 VGGNAMAVALFEFTPTFFISDKVANSVRTSFFWDMGTWDTNWDSSQSYGPDYSDP-- 763
Qy 750 TAEQNAKAANRPLLTQDKQLRYASAGVATWTPITGPIGLSISYAKPLNKKONDOTDTVQFOI 809
Db 764 -----SNIRMSAGIALQWMSPLGLPVSFAQPKKYDGDKAEOFQENI 806
Qy 810 GSVF 813
Db 807 GKTW 810

RESULT 7
B82099
surface antigen VC2252 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: B82099
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
Chardson, D.; Ermolaeva, M.D.; Vanathavan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, H.
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833
A:Accession: B82099
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-803 <HEI>
A:Cross-references: GB:AE004297; GB:AE003852; NID:9656810; PIDN:AAF95396.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC2252
A:Map position: 1
C:Superfamily: protective surface antigen D-15

Query Match 28.6%; Score 1203; DB 2; Length 803;
Best Local Similarity 33.4%; Pred. No. 9.Be-62;
Matches 278; Conservative 155; Mismatches 328; Indels 72; Gaps 15;

Qy 16 MAVVMVMSHTAAQAD-FMANDITITGLQRTVIESLQSVLPRLGQVSVENQADGVKALY 74
Db 8 LATLATSVSANGAEKFFVQDIQDGLGVALGAALMLKVRVGVSDVSDQVANIIRALY 67
Qy 75 ATGNFSDVQVYHQEGRIYQVTERPLIAEFNFGNRLIPKEGLQGLKNAGLAVGQPLKQ 134
Db 68 SSGNFEDVKVLRDGNLTLMVQVKERTIASVFSNKAIKEQLKQNLKQNLKQNLKQNLKQ 127
Qy 135 ATVOMIETELTNQVISOYYNTEITVQTMLDGNRVLKDMFTFAGKPARVVDIINQNH 194
Db 128 TTLSNIEKLEGLDFYISYGVKNATKAVVVTPLPRNRADLKFVFTGVSAAKIQINFINQV 187
Qy 195 FSDADLLDVLAIKDNKI--NPLSKADRYTOEKLVTSLNENRAKYNLAGVRFELKADKLN 252
Db 188 FSDBELSRFNLNVDDVAWNNFLAD-DKTKQVLAGDIEALRYTYLDRGYLKFQVDSTQVA 246
Qy 253 INEDKNRIFVSLHEGEQYRFQGTQFLGNLTQVQAELEALLKFKAEFGFSQAMLEQTTN 312
Db 247 ISPDKGVYITLNLNEGEPTVSKVQFGLMGKAEFTSLIPEIGETNGNSAVTRLEE 306
```

```
Qy 313 NISTKFGDDGYYIAQIRPVTRINDESRTVDVEYYIDPVHVPVYVRRINFTNFKTKQDEVLR 372
Db 307 SVKKVLGSGYAYPOVRTIPEFDDEKQCVSLVHVHVEAGKRVYVROIRFVGNNSRDEVLR 366
Qy 373 REMQLEGALASNQKIQLSRARLMTGFFKHVTVTRPVNPSPOQVDVNFVVEEQSGSS 432
Db 367 REMQMEGSLWNLKSDIETGKTRNLRLGFFETVVEQTVRVPGSEDQVDLVYSYKEANSNV 426
Qy 433 TIAAGYSGSGVTFQFDYVSQNNFMGTGKHVNASFSRSETREVYSLGTMNPNPFTVNGVSOS 492
Db 427 NFGVGYGTESGVQVGLQDNFGLSGNRVGNAMINDYQKNLTLEYRDPYWNLDGVSIG 486
Qy 493 LSGYYRKTKYDNKNISNTVLDSSYSGSLSYGYPIDNQRISEGLNADNTKLHGGRPMGTSN 552
Db 487 GKVEYNQFEASEAGIVDTNESYGTSLTWGFPDFELNRFEGIGYTHNKI--GNLTPIYLQ 544
Qy 553 VKQLMA-----DGGKIQVDNNGIPDPKHDYTTYNAILGNWYSSLDPRVPTQMSHS- 604
Db 545 VENFLAAQASNIDSGNLLTDD-----FDINLSWTRNRLNNSYFPPTAG-NHQ 591
Qy 605 --VDLTGFGDKTHQKVYVYQNIYRPFIKKS-----VLRGYAKLGYGN-----LP 648
Db 592 AFYKMTVPGSQAQYFKLQYDVQYFPLTKKHEFTLLRG--RLGYNGYGGQTDGKDNLEP 649
Qy 649 FYENFYAGGYSGVRGQDSSLGPRS--QAYLTARRGQOQTTLGEVVGGNALATFGSELILP 706
Db 650 FYENFYAGGFTSLRGFGSGAGPRAVRYDYSNGNSDTATDDSVGNAIALASVELIIVP 709
Qy 707 LPFKGDWI-DOVRPVIFIEGGQVFTTGMKDKOTID-----LTQFKDPQATAEQNAKANR 760
Db 710 TPFASEARNQIRTSIFYDMAVMDTEFDYRGKADYGNQYYVDSDP----- 756
Qy 761 PLLTQDKQLRYASAGVATWTPITGPIGLSISYAKPLNKKONDQTDVTFQIGVSFV 813
Db 757 -----TNRSYSGVALQWSPGMLPVSFLAKPIKKYEGDDEEFTTIGRTF 803

RESULT 8
B82731
outer membrane antigen XFL046 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C:Accession: B82731
R:anonymous, The xylella fastidiosa Consortium of the Organization for Nucleotide Seq
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen xylella fastidiosa.
A:Reference number: A82515; MUID:20365717
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: B82731
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-784 <STM>
A:Cross-references: GB:AE003941; GB:AE003849; NID:9105978; PIDN:AAF83856.1; GSPDB:GN
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre
as-Neto, E.; Docena, C.; El-Dorriy, H.; Facincani, A.P.; Ferrelira, A.J.S.
submitted to Genbank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Saw
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XFL046
C:Superfamily: protective surface antigen D-15
```

Query Match 27.9%; Score 1172.5; DB 2; Length 784;  
Best Local Similarity 33.8%; Pred. No. 5.5e-60;  
Matches 278; Conservative 151; Mismatches 321; Indels 73; Gaps 13;

QY 14 MTAVMVMVSTHAQAAD-FNANDITITGLQRTVIESLQSVLPFRGLQGVVSENQADGKA 72  
DB 12 LTAANFSLPVLOAAESFVANDIRVDGLQRIASGVFTYLPVNRGDTVDKADVAIRA 71  
QY 73 LYATGFSQVQVYHQEGRIIVQTERPLIAEINFEGRNRLIPKEGLQEGKNAGLAVGQPL 132  
DB 72 LYRTGFENVIRDRQGNILVVKVERPAINKLITGNKDKSEBLLKGLSEIGLSEGGTF 131  
QY 133 KQATVQMIEITELNQYISQGYNTFIVKOTMLDGNRVKLDMTFAEGKPARVVDINIIGN 192  
DB 132 DRLSLDRVTOELKROYNRRKGVNQMTTTPPLDRNRVDVTIAIKEGAARIRHINLIGT 191  
QY 193 QHESDADLIDVLAIKD-NKINPLSKADRYTOEKLVTSLLENRAKYLINAGVREIKDAKL 251  
DB 192 EKENNDKVMASWESKEHNWASWYRDDQYSKEKLSGDKLNSWYLDRGVDFNIDSTQV 251  
QY 252 NINEDKNRIEVEISLHEGEQYRFQOTQFLGNLTAVYTOAELEALLKFAEEGFSQAMLEQTT 311  
DB 252 SISPEKHMFITAGVTEGDQYKISSIKVTGNTVLPQEKIEKLVIPKTDGIFSRVLLSEYS 311  
QY 312 NNISTFGDDGYYYAQIRPVTRINDESRVDEYYIDPVHPVVRINFTGNPKTQDEVL 371  
DB 312 AAILNTLSNIGYAFSKSVPNIPPTANRADRTVAVNLHVIPGRVTVRQILFKGNTRTSDEVL 371  
QY 372 RREMROLEGALASNOKIQLSRARLMTGFPKHVTVDTRPVNPSQDQVDVNFVVEEQPSGS 431  
DB 372 RREMROFENSWYSOAALDRSKIRLQRUGYFEADVDESTVPVGSNDQVDIVYTKETSGS 431  
QY 432 STIAAGYSQGGVTFOFVDSQNNFMGTGKHVNASFSRSETREYVSLGNTNPNYFTVANGVSQ 491  
DB 432 FOYGLGYSKYTVTTVOLSONFLSGNRVSDASRSRYQDRYSPSYNPNPFTDNGVSL 491  
QY 492 SLGYSKTKYDNKNKNTSNVYLDYSGSLSYGYPIDENQRIISFGLNADNTKLHGGRFMGIS 551  
DB 492 GYNLAYOKLDYDFNAAYNSKRMSTGTFIPGIPITENDTVSWVIGADSNQI--TTPPG-S 548  
QY 552 NVKOLMADGGKIQVDNNGIPDFKHDTTYNAIILGNVYSSLDPRVFPQGMHSHVDTLVGF 611  
DB 549 TPRALID-----YIDAVG-----QRTFAWTELGWARDTRNDYFMPNLGMQIRIGAEVTL 599  
QY 612 GDKT--HQKVVYQGNIRPPIKKSRLGYAKLGYGN-----LPHY 650  
DB 600 PGSTIKYKINYQISKWPIIPALVNLTRLEVGYGDDYGKSHRILPDGTIVATASGLPFF 659  
QY 651 ENFYAGYGSVRGYDQSSIGRPSQAVLTARRGOQTLTGEVVGNGALATFGSELILPLPFK 710  
DB 660 ENFYAGTNSVRGFRDNTLGRSE--VTALYNQ---GQPLGSGFKTVGSTEMYFPKLEF 713  
QY 711 GDWIDQVRPVIETEGGOVFTTGMDKQTLDTQKDPQATAEQNAKAAANRPLLTQDKOLR 770  
DB 714 S---PSARISAFIDFGVFN-----GVNFKA-----NELR 741  
QY 771 YSAGVGATWTPICPLSISYAKPLNKQNDQTDVQFQIGSVF 813  
DB 742 ASSGVALLWRAPICPISISYAFPIKKNENDEIERLQFTFGQGF 784

RESULT 9  
G81228  
outer membrane protein Omp85 NMB0182 [Imported] - Neisseria meningitidis (strain MC58 se  
C:Species: Neisseria meningitidis  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
C:Accession: G81228  
R:Rettlein, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.  
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;  
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.  
Science 287, 1809-1815, 2000  
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve  
A:title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.

A:Reference number: A81000; MUID:20175755  
A:Accession: G81228  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-797 <DET>  
A:Cross-references: GB:AE002375; GB:AE002098; NID:g7225394; PIDN:AAF40639.1; PID:g722  
A:Experimental source: serogroup B, strain MC58  
C:Genetics:  
A:Gene: NMB0182  
C:Superfamily: protective surface antigen D-15

Query Match 26.5%; Score 1115; DB 2; Length 797;  
Best Local Similarity 32.5%; Pred. No. 1.2e-56;  
Matches 266; Conservative 154; Mismatches 361; Indels 38; Gaps 14;

QY 10 QVSAMTMAVMVMVSTHAQAADFNANDITITGLQRTVIESLQSVLPFRGLQGVVSENQADG 69  
DB 2 KLKOIASALMMLGISPLALADFTIQDIRVEGLQRTSTFTVFNLPVKVGDYNDTHGSAI 61  
QY 70 VKALYATGNFSQVQVYHQEGRIIVQTERPLIAEINFEGRNRLIPKEGLQEGKNAGLAVG 129  
DB 62 IKSLEYATGFPDDVRVETADGQLLLTVIERPTIGSLNITGAKMLQNDAIKKNLESFGLAQS 121  
QY 130 QPLKQATVQMIEITELNQYISQGYNTFIVKOTMLDGNRVKLDMTFAEGKPARVVDINI 189  
DB 122 QYFNQATLNOAVAGLKEEYLGRLKNIQITPKVTKLARNRVDITITDEGSAKITDIEF 181  
QY 190 IGQHESDADLIDVLAIKDKNI-NPLSKADRYTOEKLVTSLLENRAKYLINAGVREIKD 248  
DB 182 EGNQVSDRKLMDQMSLTEGGIWTWTRSNQFNEQFAQDMKEKVTDFYQNGVDFDRILD 241  
QY 249 AKLINEDKNRIEVEISLHEGEQYRFQOTQFLGNLT-YTOAELEALLKFAEEGFSQAML 307  
DB 242 TDLOTNEDTKQIKITVHEGGRFNGKVSIEGDTNEVPKAELEKLLTMKPKWYERQOM 301  
QY 308 EQTNNISTKFGDDGYYYAQIRPVTRINDESRVDEYYIDPVHPVVRINFTGNPKTQ 367  
DB 302 TAVLGEIQNRMGSGAYSEISVQPLPNAETKTVDVFLHIEPGRKIVYNEIHTGNKTR 361  
QY 368 DEVLREMRQLEGALASNOKIQLSRARLMTGFPKHVTVDTRPVNPSQDQVDVNFVVEEQ 427  
DB 362 DEVVRELROMESAPYDTSKLQSKERVELLGYFDNVQFADVPLAGTDPKVDLNMSTER 421  
QY 428 PGSSTIAAGYSQGGVTFOFVDSQNNFMGTGKHVNASFSRSETREYVSLGNTNPNYFTVN 487  
DB 422 STGSLDLSAGWQDTGLVMSAGVSQDNLCFSGKSAALRASRSKTLNGLSLSTDPYFTAD 481  
QY 488 GVSQSLSGYYR--KTKYDNKNISNYLDYSGSLSYGYPIDENQRIISFGLNADNTKLHGG 545  
DB 482 GVSGLGYDVYKAFDPRKASTSIKQYKTTTAGAGIRMSVPVTEYDRVNFGLVAEHLTVN-- 539  
QY 546 RFWGISNVKOLMADGGKIQVDNNGIPDFKHDTTYNAIILGNVYSSLDPRVFPQGMHSHV 605  
DB 540 ---TYNKAPHYADFIKKYKKTG-DGSPFGWLYKCTVSGWRNKTDALWPRGYLTGV 595  
QY 606 DLTWGF-GDK-THQKVVYQGNIRPPIKKSRLGVLR-----GYAKLGY--NNLPYENFYAG 656  
DB 596 NAEIALPGSKLQIYSATHNQTWFFPLSKTFTLMLGGEVGIAG-GYGRTEKIPPEFNFYGG 654  
QY 657 GYGSVRGYDQSSIGRPSQAVLTARRGOQTLTGEVW--GGNALATFGSELILPLPFKDWI 714  
DB 655 GLGSVRGYESTGLGPK-----VYDEYGEKISYGGNKKANVSALLFPFGAKD-A 703  
QY 715 DOVRPVFTIEGGQVFTTGMDKQTLDTQKDPQATAEQNAKAAANRPLLTQDKOLRYAS 774  
DB 704 RTVRLSLFADAGSVWDGKTDYDDSSST-----GGRVQNIYGAGNTHKSTFTNELRYSAG 758  
QY 775 VGATWTPICPLSISYAKPLNKQNDQTDVQFQIGSVF 813  
DB 759 GAVTWLSPLGPMKFSYAYPLKKKPEDEIQRFQGLGTF 797

RESULT 10  
DB2000

outer membrane protein OMP85 NMA0085 [imported] - Neisseria meningitidis (strain 22491 s  
C:Species: Neisseria meningitidis  
C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
C:Accession: DB2000  
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel  
C.; Holroyd, S.; Jagers, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,  
Nature 404, 502-506, 2000  
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491.  
A:Reference number: A81775; MUID:2022556  
A:Accession: DB2000  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-797 <PAR>  
A:Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; PIDN:CAB83401.1; PID:g737885  
A:Experimental source: serogroup A, strain 22491  
C:Genetics:  
A:Gene: omp85; NMA0085  
C:Superfamily: protective surface antigen D-15

Query Match 26.4%; Score 1111; DB 2; Length 797;  
Best Local Similarity 32.4%; Pred. No. 2e-56;  
Matches 285; Conservative 155; Mismatches 361; Indels 38; Gaps 14;

QY 10 QVSAMTMAVMVMASTHAQAADFANDITITGLQRTVIESLQSVLPFRLLGQVSVSENOIADG 69  
DB 2 KKKQIASALMVLGISPLALADFTIQDIRVEGLQRTPESTVFNVLPVKVGDYNDTHGSAI 61  
QY 70 VKALIATGNFSDVOVYHQEGRIYQVTERPLIABINEGNRLIPKQGLQGLKNAGLAVG 129  
DB 62 IKSLYATGFFDDVRVETADQGLLITVTERPTIGSLNTGAKMLQNDIAIKKNLESFGLAQ 121  
QY 130 QPLQAQTVQMLELTNQYISQGYNTETVTKQPTMLDGNRVKLDMTFAEGKPARVVDINI 189  
DB 122 QYFNQAQLNQAVALGKEEYLGKGLNLTQYKPVTKLARNRVDIDITIDEGRSAKITDIEF 181  
QY 190 IGNOHFSDADLIDVLAITKDNI-NPLSKADRYTQEKLVITSLLENRAKYLNAGVRFEEKD 248  
DB 182 EGNQVYSDRKLRMOMSLTEGGIWTWLTFSNOFNQKFAQDWEKVTDFYQNNGYDFDRILD 241  
QY 249 AKLNINEDKNRIFVEISLHGEQYRFQTOFLGNLT-YTQAELEALLKFAEGFSOAML 307  
DB 242 TDIQTNEDEKTKQTIKITVHEGGFRWGVSTEGTNEVPKAELEKLLTMKPGWYERQOM 301  
QY 308 EQTTNNISTKFGDDGYAQAIRPVTRINDESRTVDVEYIDPVHPVYVRRINFTNFEKTQ 367  
DB 302 TAVLGETONRMGSAGYAYSEISVOPLNPAETKTVDVFLHIEPGRKIYVNEIHTGNKTR 361  
QY 368 DEVLREMRLEGALASNOKITQLSRARLMRTGFFKHHVTVDTRPVPSNDVQDVNFVVEEQ 427  
DB 362 DEVVRELROMESAPYDTKLORSKERVLLGYFDNVQFQDAVPLAGTPDKVDLNMSLTER 421  
QY 428 PSGSSTIAAGYSQGGVTFQFDYSQNNPMFGTKHVNASFSEKSETPREYISLGMTNPFYTN 487  
DB 422 STGSLDLSAGVQVQDTGLVMAGVSQDNLFGTGKSAALRASRSKTTLLNGSLSFDPYETAD 481  
QY 488 GVSOSLSGYR--KTKYDKNKINSYVLDSVGGSLSYGYPIDENORISFGLNADNTKLHGG 545  
DB 482 GVSLSGYDVGKAFDPRKASTSIQYKTTAGATIRMSVPYTEYDRVNFVAEHLTV-- 539  
QY 546 RFMGISNVKOLMADGGIKQVDNNIGIPDKHDYTYNAILGNWYSSLDPRFPPTQCMSHSV 605  
DB 540 ---TYNKAPKHYADFIKYGGKTDG--TGDSFKGWLKYGTVGGRNKTDSALMPTRGYLTGV 595  
QY 606 DLTYGF-GDK--THQKVYVQGNIRYFFIKKSIVR----GYAKLGYG--NNLPFFYENFYAG 656  
DB 596 NAEIALPGSKLQYYSATHNQTWFFPLSKTFTMLLGGGVGIAG-CYGRKKEIPFENFYGG 654  
QY 657 GYGSVRGYDQSSLPGRSQAYLTARRGOQTTLGEVV--GGNALATFGSELILPLPFKGDWI 714  
DB 655 GLGSVRGYESTGLGPK-----VYDEGEKISYGGNKKANVSALLFPMPGAKD-A 703

Qy	715	DQVRPVIFEGGVFTTGMDKQTIDLTQPKDPQAETAQNKAANRPLLTQDKQLRYSG	718
		:       :     :	
Db	704	RTVRLSLFADAGSVGCKTYDDNSSAT-----GGRVNIIYGAGNTHKSTFTNELRYSAG	758
		:       :     :	
Qy	775	VGATWTPIGPISLSIAKLKKKNDQDTDVQFGISVF	813
		:       :     :	
Db	759	GAVTWLSPGLPMKFESYAYPLKKKPDEIQRFFQLGTTF	797
		:       :     :	

  

RESULT	11
F64102	protective surface antigen D-15 - Haemophilus influenzae (strain Rd KW20)
C:Species	Haemophilus influenzae
C:Date	18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 18-Sep-1999
C:Accession	F64102
R:	Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.
J:	Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.
M:	M. Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geogha-
Science	269, 496-512, 1995
A:Authors	Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith,
A:Title	Whole-genome random sequencing and assembly of Haemophilus influen-
A:Reference number	A64000; MUID:95350630
A:Accession	F64102
A>Status	nucleic acid sequence not shown; translation not shown
A:Molecule type	DNA
A:Residues	1-808 <TIGR>
A:Cross-references	GB:I42023; TIGR:HI0917
C:Superfamily	protective surface antigen D-15
C:Keywords	surface antigen

  

Query Match	25.1%	Score	1053.5;	DB 2;	Length	808;	
Best Local Similarity	31.8%;	Pred. No.	4.4e-53;				
Matches	260;	Conservative	158;	Mismatches	337;	Indels	63;
							Gaps

Query Match	25.1%;	Score 1053.5;	DB 2;	Length 808;
Best Local Similarity	31.8%;	Pred. No. 4.4e-53;		
Matches	260;	Conservative 158;	Mismatches 337;	Indels 63; Gaps
Qy	23	STHAQAADFMANDITITGLQRVTIESLQSIVLPRLGVWSENOLADGKALYATGNFSDV	82	
Db	27	TTTVFAPFAFKDLRDVGVDGLEQQIRASLPVRAGORVTDNDVANIVRSILVSGREDDV	86	
Qy	83	QVYHOGRI-IYQVTERPLIAENFEGNRILPKEGLQEGUKNAGLAGVQPLKAQTVMOMIE	141	
Db	87	KA-HQEGDVIVSVAKSIISDKIGNSIIPTEALKQNLDANGFKVGDVLIREKLNEFA	145	
Qy	142	TELINQVISGYNYETLTIVKTMLDGNRVLKDMTFAEGKPARVVVDINIIGNOHFSADLI	201	
Db	146	KSVKEHYASGRYNAVTEPIVNTLPNNRAELLIQINEDDRKAKLASLTFKGNESVSSSTLQ	205	
Qy	202	DVLAIKDNKINPLSKADRYTQEKLVTSLENDRAKYLNAGFVRFEIKDAKLININEDKNRIF	261	
Db	206	EOMELQDPDSWWKL-WGNKFEGAQEKDLSROIYLLNGYAKAQITKTDVOLNDEKTKVN	266	
Qy	262	VEISLHGEOYRFQGTOFLNLITYQAELEBALK-FKABEGFSQAMLEQOTTNNISTKFGD	320	
Db	265	VTIDVNEGLOYDLRSARIIGNLGMSAELEPLLALSALHLNDFRRSADIADVENAIIKALGE	324	
Qy	321	DGYYYAIOIPVTRINDERSRTVDVEYYIDPVHPVYVRINFNGFNKCTODEVLRREMROLEG	380	
Db	325	RGYGSATVSNVPDEDDRANKLATLVVDAGRLLTVQLRFEQNTVSADSTILRDMRQOEG	384	
Qy	381	ALASNQIKLSRARLMRTGFKKHTVDTR--PVPNSPDQVDVNFVWEQPSGSTIAAGY	439	
Db	385	TWYNSQLVELGKIRLDRGTFFE--TVENRIDTPINGSNDEVDDVYVKVKERNTGSI	444	
Qy	439	SQSGCVTFQFDVSONNFMWGCKKHYNASFSESTRREVYSLGMTNPDTFYVNGVOSLSGYR	498	
Db	443	GTESGISYQASVKQDNFLGTGAAVSIAGTKNDYGTSYNLGYTPEYFTKDGV--SLGNGNVF	500	
Qy	499	KTKKYDNK--NISNWLDTSGGSLSYCPIDEHQIRISFGL----NADNTKLHGGRPMGI	550	
Db	501	FENTDNSKDSOTSSNKPTTGSNNVTLGFPVNNNSSYVYGCHTYNKSINFALEYNRNLXI	560	
Qy	551	SNVKQLMADGKGIOVDNNGIPDEFKHDTYTTHAILGWNYSSLDRPFVFTQGMHSVD--LT	608	

Db 561 QSMK-----FKNGIKTNDFFS-----FGWYNLSNRGYFPTKGKASLGGRVT 605  
Qy 609 VFGDQTHQKVVYQGNIRPF-----IKKSVLRGYAKLGYGN-NLPFFYENFYAGGYGS 660  
Db 606 IPGSDNKKYKLSADVOGFYPLDRDLHWVSASAKAGYAN-GFGNKRLPFYQTYTAGGIGS 664  
Qy 661 VRGYDQSSLSGRPSQAYLTARGQOT---TIGEVVGGNALATFGSELILPLPFKGDWI-DQ 716  
Db 665 LRGFAYGSIGPNA-IYAEHNGNGTFKKISSDVIGGNAITATASAEILVPTPFVSDKSQNT 723  
Qy 717 VRPFVIFEGGVFDTT-GMDKQITDLTQFKDPOATAEONAKAANRPLLTQDKQLRYSAGV 775  
Db 724 VRTSLFVDAASVWNTKWKSDKSLD-----NNVLKSLPDYGGKSSRIRASTGV 770  
Qy 776 GATWYTPIGPLSISYAKPLNKKQNDQTDVTQFQIGSVF 813  
Db 771 GFQWQSPIGLVFSYAKPIKKYENDVDEQFQFISIGGSF 808  
RESULT 12  
JC4078  
protective surface antigen D-15 precursor - Haemophilus influenzae (type b)  
C:Species: Haemophilus influenzae  
A:Variety: type b  
C:Date: 30-Jun-1995 #sequence\_revision 14-Jul-1995 #text\_change 29-Sep-1999  
R:Flack, F.S.; Loomore, S.; Chong, P.; Thomas, W.R.  
Gene 156, 97-99, 1995  
A:Title: The sequencing of the 80-kDa D15 protective surface antigen of Haemophilus influenzae  
A:Reference number: JC4078; MUID:95255676  
A:Accession: JC4078  
A:Molecule type: DNA  
A:Residues: 1-797 <FLA>  
A:Cross-references: GB:U13961; NID:g537447; PIDN:AAA85645.1; PID:g537448  
A:Experimental source: type b  
C:Superfamily: protective surface antigen D-15  
C:Keywords: surface antigen  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-797/Product: protective surface antigen D-15 #status predicted <MAT>  
Query Match 25.08; Score 1052.5; DB 2; Length 797;  
Best Local Similarity 31.68; Pred. No. 4.9e-53;  
Matches 260; Conservative 160; Mismatches 333; Indels 69; Gaps 20;  
Qy 23 STHQAADFMANDITITGLQRTVIESLQSVLPFRIGQVVSLENLADGVKALYATGNFSDV 82  
Db 14 TTTVFAAPFYAKDIRVDGVQDLEQQIRASLPVRAGQRTDNDVANIVRSLFVSGREDV 73  
Qy 83 QVYHOGRI-IYQVTERPLIAEINFEGRNRLIPKEGLQKNAGLAVGQPLKQATVQOMIE 141  
Db 74 KA-HQEGDLVLSVVAKSIISDVKIKGNSVIPTALKONLDANGFKYGDVLIREKLNEFA 132  
Qy 142 TELTNQYISQGYNTIETVKTMLDGNRVKLDMTFAECKPARVVDINLIGNQHSDDALI 201  
Db 133 KSVKEHVASVGRYNATVEPIVNTLPNNRAEILIQINEDDKAKLASLTFKGNESVSSTLQ 192  
Qy 202 DVLAIKONKINPLSKADRYTOEKLVTSLENLRKALYAGFVRFEIKOAKLNINEDKNRIF 261  
Db 193 EQMELQPSWKKL-WGNKFEGAQAEKDLQSIIRDYLLNGYAKAOITKTDVQLNDEKTKVN 251  
Qy 262 VEISLHEGEQYRFQGTQFLGNLTQTALEALLK-FXAEFGFSQAMLEQTTNNISTKFGD 320  
Db 252 VTIDVNEGLQDLARSARTIIGNLGMSEAELPLLSALHLNDTFRSSADIADVENAIKALIGE 311  
Qy 321 DGYYYAQIRPVRINDESRITVDVEYIDPVHPVVRINFTCNFKTODEVLRRERWOLEG 380  
Db 312 RYGSATVNSVPDDEANKTKLTLVWDAGRRLLTVROLRFEFGNTVSADSTLRQEMROEG 371  
Qy 381 ALASNQIKQLSRARLMRTGFFKHHVTVDTR--PVPNSPDVDVNFVWEQPSGSSITIAAGY 438  
Db 372 TWYNSQLVELGKIRLDRTGFFE--TVENRIDPINGSNDEVDVYKVRNRTGINSFNGIGY 429

Qy 439 SQSGGVTFQFDVSONNFMGTGKHYNASFSRSETREVYSLGMNTNPYFTVNGVSQSLSGYR 498  
Db 430 GTESGISYQASVKQDNFLGTGAASVIACTKNDYGTSVNLGYTEPYFTKDGV--SLGGNVF 487  
Qy 499 KTKYDNK---NISNVLDSYCGSLSYGYPIDENQRISFGL-----NADNTKLHGREFMGI 550  
Db 488 FENTDNKSQTSNRYKRTTYGNTYTLGFPVNNNSYVYGLGHTYNTKISNFALEYNRNLYI 547  
Qy 551 SNVKOLMADGGKIQVDNNGIPDEFKHDYTYNAILGNWYSSLDROPVFPQOGSHSVYD--LT 608  
Db 548 QSMK-----FKNGIKTNDFFS-----FGWYNLSNRGYFPTKGKASLGGRVT 592  
Qy 609 VFGDQTHQKVVYQGNIRPF-----IKKSVLRGYAKLGYGN-NLPFFYENFYAGGYGS 660  
Db 593 IPGSDNKKYKLSADVOGFYPLDRDLHWVSASAKAGYAN-GFGNKRLPFYQTYTAGGIGS 651  
Qy 661 VRGYDQSSLSGRPSQAYLTARGQOTTLG-----EVVGGNALATFGSELILPLPFKGDW 713  
Db 652 LRGFAYGSIGPNA-IYAEYNGSGSGTGFKKISSDVIGGNAITATASAEILVPTPFVSDK 708  
Qy 714 I-DOVRPVIETEGGVFDTT-GMDKQITDLTQFKDPOATAEONAKAANRPLLTQDKQLRY 771  
Db 709 SONTVRSLSFVDAASVWNTKWKSDKNGLESVLEK-----RLPDYGGKSSRIRA 755  
Qy 772 SAGVGATWYTPIGPLSISYAKPLNKKQNDQTDVTQFQIGSVF 813  
Db 756 STGVGFQWQSPIGLVFSYAKPIKKYENDVDEQFQFISIGGSF 797  
RESULT 13  
D71726  
outer membrane protein Omp1 (omp1) RPI60 - Rickettsia prowazekii  
C:Species: Rickettsia prowazekii  
C:Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 03-Nov-2000  
A:Accession: D71726  
R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark  
Nature 396, 133-140, 1998  
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.  
A:Reference number: A71630; MUID:99039499  
A:Accession: D71726  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-768 <AND>  
A:Cross-references: GB:AJ235270; GB:AJ235269; NID:g3860572; PIDN:CAA14627.1; PID:g386  
A:Experimental source: strain Madrid E  
C:Genetics:  
A:Gene: omp1; RPI60  
C:Superfamily: protective surface antigen D-15  
Query Match 17.98; Score 754; DB 2; Length 768;  
Best Local Similarity 26.38; Pred. No. 8.4e-36;  
Matches 216; Conservative 145; Mismatches 383; Indels 76; Gaps 20;  
Qy 7 KGFQVSAMTMAVMVMSTHQAADFMANDITITGLQRTVIESLQSVLPFRIGQVVSLENL 66  
Db 2 KILSISKLTILLTIFYYHAFADYVIRKTIIEGHNHRVERSTIESYLSKLVNVTYNNKS 61  
Qy 67 ADGVKALYATGNFSDVQVY-HQEGRIYQVTERPLIAEINFEGRNRLIPKEGLQKNAG 125  
Db 62 DEAKRLRYATSLFRINNIYTNDGNLIVNVTTETPFISSVWFSGNSKITNLAKIETWMS 121  
Qy 126 LAVGQPLKQATVQMIETELTNQYISQGYNTIETVKTQPMLDGNRVKLDMTFAECKPARV 185  
Db 122 --GESLSQAKIELDVKKILEIYKRSGRFSTKVPKIKSLNNRVKVFIDIAEGPKTVIK 178  
Qy 186 DINIIGNQHSADLIDVLAIKONK-INPLSKADRYTOEKLVTSLENLRKALYAGFVR 244  
Db 179 SIYFSGNEHYSDSELKSVLTKESEWFRFLESNDTDPDRVEYDKELREFYQSVGFAD 238  
Qy 245 EIKAKLNINEDKNRIFVEISLHEGEQYRFQGTQFLGNLTQTALEALLKFAEALKFAE 303

Db 239 RVISASVALNDKEVFTTYSIEBEKEYRFGNVTIDNKLNTINIKOLANKVINIKOGKIFN 298  
Qy 304 OAMLEPQTNNISTKFGDDGYIAQRPVTRINDESRVTDVEXYIDPVHPVYVRRINFTGN 363  
Db 299 MKTVDDIAEKIGEYTAGYPAVNVYDIIKND-NHTADIKFIIEKADKVKINKINIINN 357  
Qy 364 FKTDQEVLRMRQLEGALASNQKIQLSRARLMRTGFFKHVTVDRPVPNSPDQVDVNFV 423  
Db 358 LKTEDHVIRRAFKIEGDVMNRSYTEKGERNLRLNDYFEKVSISLAQT-KAKDKYDVNVE 416  
Qy 424 VEEQPSGSTTAAGYSQSGVTFQDVSONNFMGTGKHVNASFSETSRETVYSLGTMNYP 483  
Db 417 VDEKSTSIGFDLGYNTAGLGRFSFLERNLVGFKILLNAGVQVSKNSTYGCITDPH 476  
Qy 484 FTVNGVSQSLSGY----RKTNDKNKINSVLDVSGSLSGVGPIDE--NORISFGLNA 537  
Db 477 FLDRDLISLVNAFRNYTRGASVLTNTDQSYKLHSGIKVIGYDMKEDLSHEIDYLKR 536  
Qy 538 DNTKLHGFRMGISNVKOLMADGGKIQVDNNGIPDKHDTYTYNAILGNWYSSSLDRPVP 597  
Db 537 DILSAPSP-----SNSIFLNEQMGKLIITSAIG-----HTIT-----VDQTDNKIVP 577  
Qy 598 TQG--MHSVDLTVGFGDKTHOKVYVYOGNIYRPFITKSVLRGYAKLG-----YGNLNP 649  
Db 578 KNGYLVSTQEFAGVGGDKYIKHIECKFYKSFINNKTLSAAGDMAGLGKGMVRI 637  
Qy 650 YENFYAGGYGVRGYSQSLGPRSOAYLTARRGOQTTLGEVVGGNALATFGSELILPLPF 709  
Db 638 SDRFNLDY-SURFASGVGPR-----EKNTEGLGGERYTFSTELNFFTPV 685  
Qy 710 KGDWIDQVRPFIIEGGQVDTTGMKQOTIDLTQFKDQATAEQNAKANRPLLTQDKQL 769  
Db 686 PEEF--NFTGAVFIDLGSVWG-VGLNKK-----QYKTPNG-----FYNDQSL 724  
Qy 770 RYSGAGVATWTPIGPLSISYAKPLNKKQNDQTDTRVQFI 809  
Db 725 RASVGFGIWIWTRFAPIRMDWGFPIKKQYD--DTQNEHL 762

RESULT 14  
D97527  
ompl protein precursor (U51683) [imported] - Agrobacterium tumefaciens (strain C58, Cere  
C:Species: Agrobacterium tumefaciens  
C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 11-Jan-2002  
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
A:Reference number: A97359; PMID:11743194  
A:Accession: D97527  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-774 <KUR>  
A:Cross-references: GB:AE007869; PIDN:AAK87173.1; PID:g15156447; GSPDB:GN00169  
C:Genetics:  
A:Gene: AGR\_C\_2554  
A:Map position: circular chromosome

Query Match 17.5%; Score 734; DB 2; Length 774;  
Best Local Similarity 26.4%; Pred. No. 1.2e-34;  
Matches 219; Conservative 150; Mismatches 354; Indels 106; Gaps 22;

Qy 15 TMAYMVMTSTHAQADFNANDITITGLORVTIESLQSVLPRLGQVYVSENLADGVKALY 74  
Db 22 SVAGLGLVASGAVNAAVISIDVRGABRSAGDSVRSNITITAPKNFSNDSIDESVKRLY 81  
Qy 75 ATGNFSDQVYHQEGRIITYQVTRPLIAEINFEGNRLPKEGLQEGLNAGLAVGQPL-- 132  
Db 82 ATGYFSNVMSRVSGSLVWTVYNEQNLVQNVVFNENRKRKDD-----KLAGIVQTQMPG 135  
Qy 133 -KQATVQMIELTANQYISQYGYNTEITVTKQTMLDGNRVRKLDMTFAEGKPARVVVDINIIG 191

Db 136 FNOAIVTADIARIKEAYSAGIRSDVEITTTQTVSVGGVRNIAFVINEGERTKIGRIDFIG 195  
Qy 192 NOHFSADLIDLVL-AIKONKINPLSKADRYTOEKLVTSLNLRKAYLNAGFVRFPEIKDAK 250  
Db 196 NNSYSDGLAALAVINTPKSNMLSFLTRKDYNEDKLRADDEALRQFYNNRGYADFVYSSD 255  
Qy 251 LNIINDEKNRIFVEISLHEGEQYRFQOTQFLGNLTVTO-AELEALLKFAAEFGSQAMLBO 309  
Db 256 AVLDESKNEYIISITVDEGKKYDFGNVAVESTVPGVDSGSELQGLVETROGASYSAKEVOQ 315  
Qy 310 TTNNTSKFGDDGYIAQIRPVTRINDESRVTDVEXYIDPVHPVYVRRINFTGNFKTQDE 369  
Db 316 SMEALSKRVAEGYFARVTPRGDRDMSGNTIGVTYIVDQGERAYVERIEIRGNTRTRDY 375  
Qy 370 VLRRMRQLEGALASNQIKIQLSRARLMRTGFFKHVTVDRPVPNSPDQVDVNFVVEEQPS 429  
Db 376 VIRREFDISGDAFNQTIITAAKRRLALGYESFYKNISITAG-GSAPDRVYIVVDVEDQST 434  
Qy 430 GSSTIAAGYSQSGVTFQFDVSONNFMGTGKHVN-ASFSSRSETREVYSLGTMNPTFTVNG 488  
Db 435 GSFGIGAGYSQNDGVLLEASVEEKNFLGRGQYIRVAAGAGEDDARTYLSLSTPEYF---- 490  
Qy 489 VSQSLSGYYRKTNDKNISNVLDVSGSLSGYYPIDENQIRISFGLNADNTKLHGGPRM 548  
Db 491 LGYRLAAGFDLTKNQSKEDEYNYDEQGFALRVATAPITENLSLTFKY----- 537  
Qy 549 GISNVKQLMADGGKIQVDNNGIPDKHDTYTYNAIL---GW-----NYSSLDPRVP 596  
Db 538 ---TYKQINYEKGQDWQNNALAE-----PYQALIRGEDMTQSLSTLNTLNTLDRNM 588  
Qy 597 PTQGMHSHVDLTVGP---GDKTHOKVYVYQVNIYRPFITK---SVLRGYA--KLGYNN 646  
Db 589 PREGHQAA--LTNEFAGLGGDSEYKIKAKARYTYLSDYDVIGSLGQAGHVMTGDN 646  
Qy 647 LPFYENFYAGGYGVRGYDQSLGPRSOAYLTARRGOQTTLG-EYVGGNALTFGSELIL 705  
Db 647 LLVFDQFKFEGG-RQVRGEKNDGIGPR-----IGSDSIGGTTTYFAASAENVTA 691  
Qy 706 PLP-FKGDWIDQVRPVIIEGGQVFDTTGMKQOTIDLTQFKDQATAEQNAKANRPLLT 764  
Db 692 PMPGVPEDF--GLRLAGFVDAGTMTYGNKVSQTV-----KD----- 726  
Qy 765 QDKQLRYSGAGVATWTPIGPLSISYAKPLNKKQNDQTDTRVQFQIGSVF 813  
Db 727 -DNSIRASAGIGVMWASFPPIRVDYAIPAKEDYDEQORFEGMSNTF 774

RESULT 15  
AE2746  
group 1 outer membrane protein precursor ompl [imported] - Agrobacterium tumefaciens  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002  
C:Accession: AE2746  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.B.; Chen, Y.; Woo  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavain, T.; Levy, R.; Li, M.; McCI  
: Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam  
ster, E.W.  
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AB2577; PMID:11743193  
A:Accession: AE2746  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-774 <KUR>  
A:Cross-references: GB:AE008688; PIDN:AAK42387.1; PID:g17739796; GSPDB:GN00186  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: ompl  
A:Map position: circular chromosome

```
Query Match      17.5%; Score:734; DB 2; Length 774;
Best Local Similarity 26.4%; Pred. No.1.2e-34;
Matches 219; Conservative 150; Mismatches 354; Indels 106; Gaps 22;

QY 15 TMAYVMVSTHAQAADFMANDITITGLQRYTIESLQSVLPFRLGQVVSENLADGVKALY 74
Db 22 SVAGLGVLASAGVANAAYISKIDVGRAGERSGADSVRSNITITAPGNFNSDIDIESVVRLY 81
QY 75 ATGNFSDQVYHQBGRILYQVTERPLIAEFNFEGRNLIPKEGLQEGLNAGLAVGQPL-- 132
Db 82 ATGYFSNVSMRVSGSTLVVTVNENOLNVQVFNENRKIKDD-----KLAGIVQTQPMGP 135
QY 133 -KQNTVMITELTNQVISOGYYNTEITVQTMLDGNRVKLDMTFAGKPARVVDINIIG 191
Db 136 FNOAIVTADIARIKEAYSAGRSDEVEITQTTSVSGQGRVNIATFVINEGERTKIGRIDFIG 195
QY 192 NOHFSADLLIDL-AIKDNKINPLSKADRYTOEKLVTSLNLRKAKYLNAGFVRFEIKDAK 250
Db 196 NNSYSDGRLAIVINTKSNMLSFLTRKDVYNEOKLRADEALROFYNNRGIAFRVYSSD 255
QY 251 LNTNEDKNRIFVEISLHEGEQYRFGQTFGLNLTQO-AELEALLKFKABEGFSQAMLEQ 309
Db 256 AVLDESKNEVTISITVDEGKKYDFGNVAVESTVPGVDGSELOGLVETROGASYSACEVQQ 315
QY 310 TTNNISTKFGDDGYYYAQIRPVTINDESRVDVEYYIDPVHPVYVRRINFTGNFKTQDE 369
Db 316 SMEAISKRVAAGEGYPFARVTPRGDRDSMGNTIGVTYIVDQGERAYVERIEIRNGTRTRY 375
QY 370 VLREMRQLEGALASNOKIOLSRARLMRTGFFKHVTYDTRPVNPSDPQDVNFVVEBOPS 429
Db 376 VIRREFDISGDAPNQTIIIAAKRRLEALGYFSKVNISTAG-GSAPDRVIVVDVEDQST 434
QY 430 GSSTIAAGYSQSGGVTFQFVDSQNNFMGTGKHVN-ASFRRSETREVYSLGMTNPYFTVNG 488
Db 435 GSGGIGAGYQNDGVLLEASVEEKNFLRGQYIRVAAGAGEDDARTYSLSTPEYF--- 490
QY 489 VSQSLSGYYRKTVDNKNISNYVLDSYGGSLSGYPIDENQRIQSFGLNADNTKLHGGRFM 548
Db 491 LGYRLAAGFDLKNQSKSEDIYNYDEQGFALRVTAITENLSTTFKY----- 537
QY 549 GISNVKQIMADGGKIQVDNNGIPDFKHDTYNAIL---GW-----NYSLSLRPVF 596
Db 538 ---TYKQINYEKGQDQWNNANLAE-----PYQALIRGEDWTQSILSNTLNYNTLDDRN 588
QY 597 PTQGMSHSVDLTVGF---GDKTHQKVYQGNIRYRPFIKK-----SVLRGYA--KLGYGNN 646
Db 589 PREGWQAA--LTNEFAGLGGDEYKIKYAKARYYYTSLDEYDVIGSLTGQAGHVMPTGDN 646
QY 647 LPFYENFYAGGYGVRGYDQSSLLGPRSOAYLTARRGQOITLG-EVVGGNALATFGSELIL 705
Db 647 LLVFDQFKFGG-RQVRGFKNDGIGPR-----IGSDSIGGTTYFAASAEVTA 691
QY 706 PLP-FKGDWIDQVRPIFIEGGQVFDTTGMDKQITIDLQFKDPOATAEQNAKAANRPLLT 764
Db 692 PMPGVPEDF--GLRAGFVDAGTMYGNKVSTSQIV-----KD----- 726
QY 765 QDKQLRYAGVGATWYTPIGPLSISYAKPLNKKONQDQTDVQFQIGSVF 813
Db 727 -DNSIRASAGIGVWWSPPFGPIRVDAIPAIAKEDYDEEQRFRCGMSNTF 774
```

Search completed: September 5, 2002, 09:55:54  
Job time: 273 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 09:54:47 : Search time 36.65 Seconds  
(without alignments)  
858.908 Million cell updates/sec

Title: us-09-701-711-2

Perfect score: 4202

Sequence: 1 MRNSYFKGFQVSAMTAVMM.....LNKKQNDQDTVQFQIGSVF 813

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1210	28.8	810	1	UP05_ECOLI
2	1053.5	25.1	795	1	D152_HAEIN
3	1052.5	25.0	793	1	D153_HAEIN
4	1052.5	25.0	797	1	D151_HAEIN
5	641	15.3	617	1	P237_BUCAI
6	256.5	6.1	578	1	YTFM_HAEIN
7	227	5.4	577	1	YTFM_ECOLI
8	197.5	4.7	469	1	CG51_HUMAN
9	171	4.1	463	1	U140_DROME
10	157.5	3.7	1577	1	HLVA_PROMI
11	134.5	3.7	1256	1	MRP_STRSU
12	149	3.5	1902	1	P3P_LACLC
13	148	3.5	475	1	YDF6_SCHPO
14	146	3.5	826	1	YEH8_ECOLI
15	146	3.5	1902	1	P1P_LACLC
16	143	3.4	1902	1	P2P_LACPA
17	142.5	3.4	705	1	PPCF_FLAME
18	140.5	3.3	1902	1	P2P_LACLC
19	140	3.3	2201	1	TENA_HUMAN
20	139.5	3.3	1324	1	CUT3_SCHPO
21	137.5	3.3	735	1	FCT_ERWCH
22	136	3.2	917	1	HXA3_HAEIN
23	135.5	3.2	1332	1	XKDO_BAGSU
24	134.5	3.2	866	1	YCBS_ECOLI
25	134.5	3.2	1178	1	CIAC_BACTK
26	134	3.2	655	1	YKDA_MYCCA
27	133.5	3.2	937	1	CS32_ECOLI
28	133	3.2	398	1	YLIB_CAEEL
29	133	3.2	1048	1	AGOI_ARATH
30	132.5	3.2	863	1	YEJO_ECOLI
31	130.5	3.1	1957	1	YD86_SCHPO
32	128.5	3.1	905	1	HXA1_HAEIN
33	127.5	3.0	839	1	YDDB_HAEIN
34	127	3.0	484	1	YNC6_YEAST
35	127	3.0	922	1	DF01_RICFE
36	126.5	3.0	1258	1	ICEN_ERWHE
37	126	3.0	925	1	VPH_BPHP1
38	126	3.0	1034	1	ICEN_PANAN
39	125.5	3.0	880	1	LYTD_BACSU
40	125.5	3.0	2022	1	ANTI_ONCVO
41	124.5	3.0	1341	1	VG37_BPT2
42	124.5	3.0	1754	1	PWPB_CHLFR
43	124.5	3.0	2366	1	PTXB_CLODI
44	124	3.0	1630	1	MSPL_PLAFK
45	124	3.0	1639	1	MSPL_PLAFW

#### ALIGNMENTS

RESULT 1

UP05\_ECOLI  
ID UP05\_ECOLI STANDARD: PRT: 810 AA.  
AC P39170; P39181; P77465;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Unknown protein from 2D-page spots M62/M63/O3/O9/T35 precursor.  
GN YAEI OR B0177 OR Z0188 OR ECS0179.  
OS Escherichia coli, and.  
OS Escherichia coli O157:H7.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562, 83334;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12";  
RL Science 277:1453-1474(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,  
RA Davis K., Federspiel N., Hyman R., Kallman S., Komp C., Kurdi O.,  
RA Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,  
RA Davis R.W.;  
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
RX MEDLINE=21074935; PubMed=11206551;  
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
RA Postai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,  
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
RA Welch R.A., Blattner F.R.;  
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";  
RL Nature 409:529-533(2001).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / RIMD 0509952;  
RX MEDLINE=21156231; PubMed=11258796;  
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,  
RA Kubara S., Shiba T., Hattori M., Shinagawa H.;  
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
O157:H7 and genomic comparison with a laboratory strain K-12";  
RL DNA Res. 8:11-22(2001).  
RN [5]  
RP SEQUENCE OF 21-32 AND 351-362.

P53969 saccharomyc  
Q9raa9 rickettsia  
P16239 erwinia her  
P51735 bacterioph  
Q47879 pantoea ana  
P39848 bacillus su  
P21249 onchocerca  
P07067 bacterioph  
O84418 chlamydia t  
P18177 clostridium  
P04932 plasmodium  
P04933 plasmodium







```

Query Match      25.1%; Score 1053.5; DB 1; Length 795;
Best Local Similarity 31.8%; Pred. No. 1.9e-54;
Matches 260; Conservative 158; Mismatches 337; Indels 63; Gaps 20;

QY 23 STHAAQADFMANDITITGLQRTVIESLQSVLPFRIGQVYSENQADGVKALYATGNFSDV 82
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 14 TTTVAAPFPAKDIRVGVQDLEQIIRASLPVRAQRTDNDVANIIVRSFVSGRFDV 73
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 83 QVYHOGRI-IQVTERPLIAEINFEGRNLPKEGLQEGKLNAGLAVGQPLKQATVQMI 141
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 74 KA-HOEGDVLVSVVAKSIISDVKIGNSIIPTEALKQNLNDANGPKFQDVLIREKNEFA 132
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 142 TELTNOYISQGYNTETITVQTMLDGNRVKLDMTFAEGKPARVDINIIGNHFSADLI 201
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 133 KSVKEHYASVGRYNATVEPIVNTLPNNRAEILIQINEDDKAKLASLTFKGNESVSSLTQ 192
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 202 DVLAIKDNKINPLSKADRYTOEKLVTSLNRAKYLNAGFVFEIKDAKLINEDKNRIF 261
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 193 EQMELQPDSSWWKL-WGNKFEQAQFEKDLQAIIRDYLLNNGYAKAQITKTDVQLNDEKTKVN 251
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 262 VESLHGEQYRFGQTOFLGNLTYYQAELEALLK-FKAEEGFSQAMLEQTTNNISTKFGD 320
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 252 VTIDVNEGLQYDLRSARIIGNLGMSAELEPLLSALHLNDTFRRSADIADVENAIKAKLGE 311
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 321 DGYVYAAQIRPVTRINDESRTVDVEYIDPVHPVYVRRINFTGNFTQDEVLRREMQL 380
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 312 RGYNTVNSVPDFDDANKTLAITFVVDAGRRLTVQRLFEQNTVSADSTLQKQEMRQEG 371
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 381 ALASNQIKQLSRARLMRTGFFKHVTVDR--PVPNSPDQVDVNFVVEQSPGSSSTIAAGY 438
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 372 TWYNSQLVELGKIRLDRGTGFFE--TVENRIDPINGSNDEVDVYVYKERTGTSINFGY 429
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 439 SOSGGVTQFQVDSQNNFMCTGKHVNASFSRSETREVYSLGMTNPNFTVNGVQSLSGYR 498
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 430 GTESGISYQTSIKQDNFLGTGAASVIACTKNDYGTGTVSNLGYTEPYFTKQGV--SLGGINF 487
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 499 KTKYDNK---NLSNVLDYSGSLSGYPIDENQISFGL-----NADNTKLHGFRPMGI 550
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 488 FENYDNKSDTSSNYKRTYGSNVTGLGFPVNNNSYVGLGHTYKINSFALEYNNLYI 547
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Query Match      25.0%; Score 1052.5; DB 1; Length 793;
Best Local Similarity 31.8%; Pred. No. 1.9e-54;
Matches 260; Conservative 160; Mismatches 333; Indels 65; Gaps 21;

QY 23 STHAAQADFMANDITITGLQRTVIESLQSVLPFRIGQVYSENQADGVKALYATGNFSDV 82
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 14 TTTVAAPFPAKDIRVGVQDLEQIIRASLPVRAQRTDNDVANIIVRSFVSGRFDV 73
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 83 QVYHOGRI-IQVTERPLIAEINFEGRNLPKEGLQEGKLNAGLAVGQPLKQATVQMI 141
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 74 KA-HOEGDVLVSVVAKSIISDVKIGNSIIPTEALKQNLNDANGPKFQDVLIREKNEFA 132
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 142 TELTNOYISQGYNTETITVQTMLDGNRVKLDMTFAEGKPARVDINIIGNHFSADLI 201
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 133 KSVKEHYASVGRYNATVEPIVNTLPNNRAEILIQINEDDKAKLASLTFKGNESVSSLTQ 192
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 202 DVLAIKDNKINPLSKADRYTOEKLVTSLNRAKYLNAGFVFEIKDAKLINEDKNRIF 261
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 193 EQMELQPDSSWWKL-WGNKFEQAQFEKDLQAIIRDYLLNNGYAKAQITKTDVQLNDEKTKVN 251
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 262 VESLHGEQYRFGQTOFLGNLTYYQAELEALLK-FKAEEGFSQAMLEQTTNNISTKFGD 320
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 252 VTIDVNEGLQYDLRSARIIGNLGMSAELEPLLSALHLNDTFRRSADIADVENAIKAKLGE 311
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 321 DGYVYAAQIRPVTRINDESRTVDVEYIDPVHPVYVRRINFTGNFTQDEVLRREMQL 380
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 312 RGYNTVNSVPDFDDANKTLAITFVVDAGRRLTVQRLFEQNTVSADSTLQKQEMRQEG 371
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 381 ALASNQIKQLSRARLMRTGFFKHVTVDR--PVPNSPDQVDVNFVVEQSPGSSSTIAAGY 438
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 372 TWYNSQLVELGKIRLDRGTGFFE--TVENRIDPINGSNDEVDVYVYKERTGTSINFGY 429
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 439 SOSGGVTQFQVDSQNNFMCTGKHVNASFSRSETREVYSLGMTNPNFTVNGVQSLSGYR 498
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 430 GTESGISYQTSIKQDNFLGTGAASVIACTKNDYGTGTVSNLGYTEPYFTKQGV--SLGGINF 487
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 499 KTKYDNK---NLSNVLDYSGSLSGYPIDENQISFGL-----NADNTKLHGFRPMGI 550
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 488 FENYDNKSDTSSNYKRTYGSNVTGLGFPVNNNSYVGLGHTYKINSFALEYNNLYI 547
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
Haemophilus.
NCBI_TaxID=727;
SEQUENCE FROM N.A.
STRAIN-PAK 12085;
MEDLINE-97427952; PubMed=9284140;
Loosmore S.M., Yang Y.P., Coleman D.C., Shortreed J.M., England D.M.,
Klein M.H.;
"Outer membrane protein D15 is conserved among Haemophilus influenzae
species and may represent a universal protective antigen against
invasive disease.";
Infect. Immun. 65:3701-3707(1997).
-!- SUBCELLULAR LOCATION: Outer membrane.
-!- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; U60834; AAB61977.1; -.
InterPro; IPR00184; Bac_surfAg_D15.
Pfam; PF01103; Bac_surface_Ag; 1.
Kw: Antigen; Outer membrane; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 793 PROTECTIVE SURFACE ANTIGEN D15.
SQ SEQUENCE 793 AA; 87511 MW; 51BFDB2036801A1A CRC64;
-----
Query Match      25.0%; Score 1052.5; DB 1; Length 793;
Best Local Similarity 31.8%; Pred. No. 1.9e-54;
Matches 260; Conservative 160; Mismatches 333; Indels 65; Gaps 21;

QY 23 STHAAQADFMANDITITGLQRTVIESLQSVLPFRIGQVYSENQADGVKALYATGNFSDV 82
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 14 TTTVAAPFPAKDIRVGVQDLEQIIRASLPVRAQRTDNDVANIIVRSFVSGRFDV 73
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 83 QVYHOGRI-IQVTERPLIAEINFEGRNLPKEGLQEGKLNAGLAVGQPLKQATVQMI 141
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 74 KA-HOEGDVLVSVVAKSIISDVKIGNSIIPTEALKQNLNDANGPKFQDVLIREKNEFA 132
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 142 TELTNOYISQGYNTETITVQTMLDGNRVKLDMTFAEGKPARVDINIIGNHFSADLI 201
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 133 KSVKEHYASVGRYNATVEPIVNTLPNNRAEILIQINEDDKAKLASLTFKGNESVSSLTQ 192
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 202 DVLAIKDNKINPLSKADRYTOEKLVTSLNRAKYLNAGFVFEIKDAKLINEDKNRIF 261
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 193 EQMELQPDSSWWKL-WGNKFEQAQFEKDLQAIIRDYLLNNGYAKAQITKTDVQLNDEKTKVN 251
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 262 VESLHGEQYRFGQTOFLGNLTYYQAELEALLK-FKAEEGFSQAMLEQTTNNISTKFGD 320
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 252 VTIDVNEGLQYDLRSARIIGNLGMSAELEPLLSALHLNDTFRRSADIADVENAIKAKLGE 311
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 321 DGYVYAAQIRPVTRINDESRTVDVEYIDPVHPVYVRRINFTGNFTQDEVLRREMQL 380
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 312 RGYNTVNSVPDFDDANKTLAITFVVDAGRRLTVQRLFEQNTVSADSTLQKQEMRQEG 371
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 381 ALASNQIKQLSRARLMRTGFFKHVTVDR--PVPNSPDQVDVNFVVEQSPGSSSTIAAGY 438
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 372 TWYNSQLVELGKIRLDRGTGFFE--TVENRIDPINGSNDEVDVYVYKERTGTSINFGY 429
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 439 SOSGGVTQFQVDSQNNFMCTGKHVNASFSRSETREVYSLGMTNPNFTVNGVQSLSGYR 498
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 430 GTESGISYQTSIKQDNFLGTGAASVIACTKNDYGTGTVSNLGYTEPYFTKQGV--SLGGINF 487
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 499 KTKYDNK---NLSNVLDYSGSLSGYPIDENQISFGL-----NADNTKLHGFRPMGI 550
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 488 FENYDNKSDTSSNYKRTYGSNVTGLGFPVNNNSYVGLGHTYKINSFALEYNNLYI 547
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 551 SNVKOLMADGGKIQVDNNGIPDKHDYTYNAILGNWYSSSLDRPVFPQGMHSHVD--IT 608
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 548 QSMK-----PKNGIKTNDPFS-----FGWNTNSLRNGYFPTKGVKASILGRVT 592
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 609 VGFGDKTHQVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 660
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 593 IPGSDNKYIKLSADVQGYPLDRDLHLVWVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 651
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 661 VRGYDQSSGLGRSQAYLTARRQOQT---TLGEVVGGNALATFGSELILPLPKGDWI-DQ 716
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 652 LRGPAYGISGPNV-IYAEHNGNGTFKFKISSDVIGGNALITASAELIVPTPEVPSQNT 710
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 717 VRPVIFSGGVFDIT--GMDQOTIDTQFKDPAFAEQNAKANRPLITQDKQLRYAGV 775
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 711 VRTSLFVDAASVWNTKWSKSGLD-----NNVLKSLPDYCKSSRIRASTGV 757
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 776 GATWYTPGLPISYAKPLNKKNDOTDTVQFQIGSVF 813
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 758 GFQWOSPGLPVFSAPIKKYENDVDVEQFQFSGGSF 795
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 3
D153_HAEIN STANDARD; PRT; 793 AA.
AC O32629;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-Ag) (Outer membrane protein D15).
OS Haemophilus influenzae.
```

```

QY 551 SNVKQLMADGGKIQVDNNGIPDFKHDYTTYNAILGNWYSSLDPRVFPFTQGMSSHVD--LT 608
Db 548 QSMK-----FKNGIKTDFDES-----FGWYNSLNRGYPFTKGVKASLGRV 592
QY 609 VFGDKTHQKV--YQG-----NIYRPFIKKSVLRGVAKLGYGN-NLPFFYENYAGYGS 660
Db 593 IPGSDNKYYKLSADVQGFYPLDRHRWWSAKASAGYAN-GFGNKRLLPFYQTYTAGGIGS 651
QY 661 VRGYDQSSLPQRSQAYLTARRGOQT---TLGEVVGGNALATFGSELILPLPFKGDWI-DQ 716
Db 652 LRGFAYGSGIPGNA---IYAEHNGTGNKISSDVIGGNAITTAELIIVPFFVSDKSQNT 708
QY 717 VRPFIETEGQVFDTT-GMDKQITDITQFKDPOQATAEQNAKAANRPLLTQDKQLRYSGAV 775
Db 709 VRTSLFYDAASVWNTWKSKDNGLESKVLKD-----LPDYGKSSRIRASTGV 755
QY 776 GATWYTPIGPLSTISYAKPLANKKQNDQDTVQFQIGSVF 813
Db 756 GFQWQSPIGPLVFSYAKPIKKYENDDVEQFQSIGGSF 797

RESULT 4
D151_HAEIN STANDARD; PRT; 797 AA.
AC P46024;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (P-15-Ag) (Outer membrane protein D15).
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SEROTYPE B;
RX MEDLINE=95255676; PubMed=7737523;
RA Flack F.S., Loosmore S., Chong P., Thomas W.R.;
RT "The sequencing of the 80-kDa D15 protective surface antigen of
RT Haemophilus influenzae."
RL Gene 156:97-99(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MINNA, AND EAGAN;
RX MEDLINE=97427952; PubMed=9284140;
RA Loosmore S.M., Yang Y.P., Coleman D.C., Shortreed J.M., England D.M.,
RA Klein M.H.;
RT "Outer membrane protein D15 is conserved among Haemophilus influenzae
RT species and may represent a universal protective antigen against
RT invasive disease."
RL Infect. Immun. 65:3701-3707(1997).
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; U13961; AAA85645.1; -
CC EMBL; U60832; AAB61974.1; -
CC EMBL; U60833; AAB61976.1; -
CC InterPro: IPR000184; Bac_surfAg_D15.
CC Pfam: PF01103; Bac_surface_Ag_1.
KW Antigen; Outer membrane; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 797 PROTECTIVE SURFACE ANTIGEN D15.
SQ SEQUENCE 797 AA; 87675 MW; 2F93DE538696AF1B CRC64;

```

```

Query Match 25.0%; Score 1052.5; DB 1; Length 797;
Best Local Similarity 31.6%; Pred. No. 1.9e-54;
Matches 260; Conservative 160; Mismatches 333; Indels 69; Gaps 20;

QY 23 STIAQAADFMADITITGLORVITIESLOSVLPRFLGVQVSENQADGVKALYATGNESDV 82
Db 14 TTTTFAAPFVAKDIRVDVGQDLEQQIRASLPVRAQGVTDNDVANIRSLFVSGRDDV 73
QY 83 QVYHQEGRI-IYQVTERPLIAEINFEGRNLIPKEGLEKLNAGLAYGOPLQKQATQVOMIE 141
Db 74 KA-HQEGDVLVSVVAKSIISDVKIKNSVIPTALKQNLNDANGFKVDVLRKLENEFA 132
QY 142 TELTNOYISOGYNTTEITVKQTLMDNRKLDMTFAEGKPARVVDIINIIGNHESDADI 201
Db 133 KSVKEHYASGVRYNATVEPIVNTLPNNKAEILIQINEDDKAKLASLTFKGNESVSSSTLQ 192
QY 202 DVLAIKDNKNPLSKADRYTQEKLVTSLENLRKAYLNAGFVRFETKDAKLINEDKNRIF 261
Db 193 EQMELQPDSSWWKL-WGNKFEQAQFKDLOSIRDYLNNGYAKAQITKTDVQNLDEKTKVN 251
QY 262 VEISLHEGEQYRFQOTQFLGNLYTQAELEALLK-FKAEEGFSQAMLEQTNTNISTKPGD 320
Db 252 VTIDVNEGLQYDLRSARLIGNLGMSAELEPLLSALHLNDFRSDIADVENAIKAKIGE 311
QY 321 DGYYIAQIRPVTRINDESRTVDVEYIIDPVHPVYVRRINFTGNFKTQDEVLRMRQLEG 380
Db 312 RYGSATVNSVPDPDDANKTLAITLVVDAGRRLTVRQLRFEQNTVSADSTLRQEMROQEG 371
QY 381 ALASNQIKLSRARLMRTGKFFKHYTVDFR--PVPNSPDQVDVNFVVEBQPSGSSTIAAGY 438
Db 372 TWYNSQLVELGKIRLDRTRGFFE--TVENRIDPINGSNDEVVDVYVKERNTGINSFNGIGY 429
QY 439 SQSGVTTFQEDVSQNNFMGTGKHVNASFSRSETREVYSLGMTNPVYFTVNGVSQSLSGYR 498
Db 430 GTESGISYQASVKQDNFLGTGAAYSIAGTKNDYGTSVNLGYTEPYFTKGV--SLGDNVF 487
QY 499 KTKYDNK---NISNVLDYSYGSLSYGYPIDENQIRISGL-----NADNTKLHGGRFNGI 550
Db 488 FENTDNKSQDTSNNYKRTYTSNVTGLPFPVNNENSYVGLGHTYKNKISNFALEYNRNLYI 547
QY 551 SNVKQLMADGGKIQVDNNGIPDFKHDYTTYNAILGNWYSSLDPRVFPFTQGMSSHVD--LT 608
Db 548 QSMK-----FKNGIKTDFDES-----FGWYNSLNRGYPFTKGVKASLGRV 592
QY 609 VFGDKTHQKVYQGNIIYRPF-----IKSVLRGAYKALGYGN-NLPFFYENYAGYGS 660
Db 593 IPGSDNKYYKLSADVQGFYPLDRHLWVWSAKASAGYAN-GFGNKRLLPFYQTYTAGGIGS 651
QY 661 VRGYDQSSLPQRSQAYLTARRGOQTTLG-----EVVGGNALATFGSELILPLPFKGDW 713
Db 652 LRGFAYGSGIPGNA---IYAEYNGSGTGTFPKIISDVIGGNAITTAELIIVTFFVSDK 708
QY 714 I-DQVRPVIFIEGGQVFDTT-GMDKQITDITQFKDPOQATAEQNAKAANRPLLTQDKQLRY 771
Db 709 SQNTVRTSLFYDAASVWNTWKSKDNGLESVDLK-----RLPDYGKSSRIRA 755
QY 772 SAGVATWYTPIGPLSTISYAKPLANKKQNDQDTVQFQIGSVF 813
Db 756 STGVGFQWQSPIGPLVFSYAKPIKKYENDDVEQFQSIGGSF 797

RESULT 5
Y237_BUCAI STANDARD; PRT; 617 AA.
AC P57331;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein BU237 precursor.
GN BU237.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
symbiotic bacterium).

```



```
Db 228 LTSDPSSNPFSSVLQPNVNHKSKTVDVLEILYP-----RK 264
Qy 375 MRLEGALASNOKIQLSRARLMRTGFKHVTVDTRPVPNSPDQVDVNFVEEQSGSTI 434
Db 265 KNAME-----L 270
Qy 435 AAGYSQSGVTFQFDVSNQPMGTGKHVNASFSRSETREVYSLGNTNPFYTVNGVS-----490
Db 271 VGFTDGGVHGQ-----IGWTPWINSRGHLSRN 301
Qy 491 -----QSLSGYRKTKYDNKNISNYLSDYSGSLSYGVYPIDENQRISFGLNADNTKL- 542
Db 302 LYLAPKQTLATYRMLPL--KNPLNYID-----FVWGGEKE-----NDTNRVL 347
Qy 543 -----HGRFPMGINSVQKLMADGGKIQVDNNGCIPDFKHDTYTNAILGNWYSSL 591
Db 348 TSLALRYWNAHGWYFG-----GLRMRYDSFTQADITDKTLLLYPTVGFTRRL 397
Qy 592 DRPVFTQG-----MSHSDVLTGFGDKTHQKVYQGNIRYRPFIRKSVLRGVAKLGYGN-- 645
Db 398 RGGSFATGWDYQKITFDLSKRILWLSSEFIRKQASSAWVRVYAENHRVVARAEIGYLHTK 457
Qy 646 ---NLFPYENYAGYGVGRYDQSSGLPRSQAYLTARRGQOTLGEVVGGNALATFQSE 702
Db 458 GIEKIPPTLRFAGGDRSVRGYGYKKIAPKNRN-----GKLVGGSRLITTSLE 505
Qy 703 LILPLPFKGDWIDQVRP---VIFTEGGQVFTDGMKDQITDLTQFKDPQATAEONAKAA 758
Db 506 ---YQYQYPNWAATFADSG-----LAADNYTA- 531
Qy 759 NRPLLTDQKQLRYSGAGYWTYPIGPLSISYAKPLNKKQNDQTDTVQFQIG 810
Db 532 -----KELRYGTGCVGRNAPVGAIKFDIATPIRDKNSK--NIQFYIG 573

RESULT 7
YTFM_ECOLI
ID YTFM_ECOLI STANDARD: PRT; 577 AA.
AC P39320;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ytfm precursor.
GN YTFM OR B4220 OR Z5831 OR ECS198.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=95334362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes.";
RL Nucleic Acids Res. 23:2105-2119(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai J., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoculis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RMD 0509952;
```

```
MedLine=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -1- SIMILARITY: STRONG, TO H. INFLUENZAE HI0698.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U14003; AAA97116.1; -;
DR EMBL; AE000493; AAC77177.1; -;
DR EMBL; AE005654; AAG59418.1; -;
DR EMBL; AP002568; BAB38621.1; -;
DR EcoGene; EG12513; ytfm
KW Hypothetical protein; Signal; Complete proteome.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 577 HYPOTHETICAL PROTEIN YTFM.
SQ SEQUENCE 577 AA; 64796 MW; 253D1D5BEA744D25 CRC64;

Query Match 5.4%; Score 227; DB 1; Length 577;
Best Local Similarity 20.5%; Pred. No. 3.5e-06;
Matches 151; Conservative 82; Mismatches 238; Indels 264; Gaps 31;

Qy 115 EGOELGNAGLAVGQPLKQATVQMIETELTNQYISQGY--NTEITVKTQMLDGNRVKL 172
Db 63 DAIREGLK-----ALGYQPTIEFDLRPPKGRQV-L 94
Qy 173 DMTFAEGKPARV--VDINIIGNHFSADLIDVLAIKONKINPLSKADRYTOEKLVTSL 230
Db 95 IAKVTPGVPLTGGTVDVLRGAR-TDRDYKLKLLTRPAIGTVLNOGYENFKSLTSLA 153
Qy 231 NLRKYLNAGFVRFEIKDAKLINEDKNRIFVEISLHGEQYRFGOTQFLGNLTYTQABL 290
Db 154 -LRKGYDFSEFTK-----AQLGIALGLHKAFWIDYNSGERYRFGHVFEG-----198
Qy 291 EALLKFAEGFSQAMLEQTNNISTKTFDDGYYAQIRPVTRINDESTRVDVEYIDPV 350
Db 199 -----SQIRDE-----YLQNL 209
Qy 351 HPVYVRINFTGNFKTQDEVLRREMRQLEGALASNQIKLSRARLMRTGFKHVTV---406
Db 210 VP-----FRGDEYESKDLAEL-----NRLSATGFWNSVVPAPQF 245
Qy 407 ---DTRPVNSPDQVDVNFVEEQSGSSTIAAGYSQSGG--VTFQFDVSNQPMGTGK 460
Db 246 DKARETKVLP-----LTGVVSPRTETIETGVGYSTVDGPRVKATKKPMWNSYGHSL 298
Qy 461 HYNASFSRSETREVYSLGM---TNP---YFTVNGVSQSLSGYRKYTKYDNKNISNVLD 514
Db 299 TTSTISAPTEQTLDFSKYMKPLKPLEQYLVQV-----GFKRTDLNDESDSTTLVAS 352
Qy 515 YGSLSYGYPIDENQRISF-----GLNADNTKLHGRFMGINSVQKLM 557
Db 353 RYWDLSUGMQORAINLRWLSLDHTQGEITNTMLTFPGVMISRTSRGG-----LM 402
Qy 558 ADGGKIQVDNNGIPDFKHDTYTNAILGNWYSSLDPRVFPPTQCMHSDVLTGVFGDKTHQ 617
Db 403 PTWGDQ-----RYSIDYSN---TAWG-SDVDFSVFQAQN-----433
Qy 618 KVVYQGNIV--RPFTKSVLRGYAKLGYGNLPLPFYENFVAGGYGVGRYDQSSGLPRSQ 675
Db 434 --VWIRTLYDRHRFVTRGTL-GWIEGTGDFKVPPLDRLFFAGGDRSIRGYKYKSIAPK---487
```

QY 676 YLTARRGQQTTLGEVVGNGALATGSELILPLPFKGDWIDQVRPVVFIEGGQVDTTGM 735  
 Db 488 YAN-----GDLKGASKLIT--GSLEYQYNTGK-----WGAVEVDSGEA 525  
 QY 736 KQTIDLTQFKDQPTAEQNAKARNRPLLTQDKQLRYSAGVATWYTPIGPLISISYAKPLN 795  
 Db 526 VSDIRRSDFKT-----GTGVGVRWESPVGPIKLDFAVEVA 560  
 QY 796 KQNDQDTDTVQFQIG 810  
 Db 561 DKDE---HGLQFYIG 572

RESULT 8  
 CG51\_HUMAN  
 ID CG51\_HUMAN STANDARD: PRT; 469 AA.  
 AC Q9V512; Q9NW85; Q9VQM9;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Protein CGI-51.  
 GN CGI51.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20057165; PubMed=10591208;  
 RA Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,  
 RA Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.,  
 RA Baguley C., Bailey J., Barlow K., Bates K.N., Beasley O., Bird C.P.,  
 RA Blakey S., Bridgeman A.M., Buck D., Burgess J., Burrill W.D.,  
 RA Burton J., Carder C., Carter N.P., Chen Y., Clark G., Clegg S.M.,  
 RA Cobley V., Cole C.G., Collier R.E., Connor R.E., Conroy D., Corby N.,  
 RA Coville G.J., Cox A.V., Davis J., Dawson E., Dharm P.D., Dockree C.,  
 RA Dodsworth S.J., Durbin R.M., Ellington A., Evans K.L., Fey J.M.,  
 RA Fleming K., French L., Garner A.A., Gilbert J.G.R., Goward M.E.,  
 RA Grafham D., Griffiths M.N., Hall C., Hall R., Hall-Tamlyn G.,  
 RA Heathcote R.W., Ho S., Holmes S., Hunt S.E., Jones M.C., Kershaw J.,  
 RA Kimberley A., King A., Laird G.K., Langford C.F., Leversha M.A.,  
 RA Lloyd C., Lloyd D.M., Martyn I.D., Mashreghi-Mohammadi M.,  
 RA Matthews L., Mccann O.T., Mcclay J., McLaren S., Mcmurray A.V.,  
 RA Milne S.A., Mortimore B.J., Odell C.N., Pavitt R., Pearce A.V.,  
 RA Pearson D., Phillimore B.J., Phillips S.H., Plumb R.W., Ramsay H.,  
 RA Ramsey Y., Rogers L., Ross M.T., Scott C.E., Sehra H.K., Skuce C.D.,  
 RA Snalley S., Smith M.L., Soderlund C., Spragon L., Steward C.A.,  
 RA Sulston J.E., Swann R.M., Vaudin M., Wall M., Wallis J.M.,  
 RA Whiteley M.N., Willey D., Williams L., Williams S., Williamson H.,  
 RA Wilmer T.E., Wilming L., Wright C.L., Hubbard T., Bentley D.R.,  
 RA Beck S., Rogers J., Shimizu N., Minoshima S., Kawasaki K., Sasaki T.,  
 RA Asakawa S., Kudoh J., Shintani A., Shibuya K., Yoshizaki Y., Aoki N.,  
 RA Mitsuuyama S., Roe B.A., Chien F., Chu L., Crabtree J., Deschamps S.,  
 RA Do A., Do T., Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S.,  
 RA Lai H., Lao H.I., Lewis J., Lewis S., Lin S.-P., Loh P., Mala J.E.,  
 RA Nguyen T., Pan H., Phan S., Qian Y., Ray L., Ren Q., Shaull S.,  
 RA Sloan D., Song L., Wang Q., Wang Y., Wang Z., White J., Willingham D.,  
 RA Wu H., Yao Z., Zhan M., Zhang G., Chisoe S., Murray J., Miller N.,  
 RA Minx P., Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H.,  
 RA Bourne S., Cordes M., Du Z., Fulton L., Goela D., Graves T.,  
 RA Hawkins J., Hinds K., Kemp K., Latreille P., Layman D., Ozersky P.,  
 RA Rohlfing T., Scheet P., Walker C., Wamsley A., Wohlmann P., Pepin K.,  
 RA Nelson J., Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R.,  
 RA Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,  
 RA Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,  
 RA Edelmann L., Kim U.J., Shizuya H., Simon M.I., Dumanaki J.P.,  
 RA Peyrard M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,  
 RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,  
 RA Khan A.S., Lane L., Tilahun Y., Wright H.;  
 RT \*the DNA sequence of human chromosome 22.\*;  
 RL Nature 402:489-495(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.

RX MEDLINE=20272150; PubMed=10810093;  
 RA Lai C.-H., Chou C.-Y., Ch'ang L.-Y., Liu C.-S., Lin W.-C.;  
 RT \*Identification of novel human genes evolutionarily conserved in  
 RT Caenorhabditis elegans by comparative proteomics.\*;  
 RL Genome Res. 10:703-713(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,  
 RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
 RA Watanashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,  
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,  
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,  
 RA Ninomiya K., Iwayanagi T.;  
 RT \*NEDO human cDNA sequencing project\*;  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE UPF0140 (CGI-51) FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; AL035398; CAB51401.1; -;  
 DR EMBL; AK001087; AAB91498.1; -;  
 DR EMBL; AF151809; AAD34046.1; -;  
 FT CONFLICT 110 110 D -> G (IN REF. 3).  
 FT CONFLICT 345 345 I -> V (IN REF. 3).  
 FT CONFLICT 368 372 WAGGL -> LGRRW (IN REF. 2).  
 SQ SEQUENCE 469 AA; 51976 MW; 4F686FB7ACB08DEF CRC64;

Query Match 4.7%; Score 197.5; DB 1; Length 469;  
 Best Local Similarity 22.0%; Pred. No. 0.00014;  
 Matches 114; Conservative 82; Mismatches 212; Indels 111; Gaps 22;

QY 334 INDESRTVDVE---YYIDPVHVVVRINFTGNFKTODEVLRREMQL---EGALASNQ 386  
 Db 22 LGEEAEFVEVEPEAKQETLENKDVVVHVHFDGLGTGDKDDIIICEIGDVFKAKNLIEVMR 81  
 QY 387 KIQLSRARLMTGTFKFKHTV--DTRPVPNS-PDQVNVFVVEE--QPSGSSTIAAGYSQS 441  
 Db 82 KSEAREKLLRLGIFRQVDVLDTCCGDDALPGLDVTFTVTELRRLTGSYNTWYGNNEG 141  
 QY 442 GGVTFQFDVSONNFMGTCKHVNASFSTRSETRVYSLGTMNP-----YFTVNGVQSLS 494  
 Db 142 SMV---LGLKLPNLIGRAEKVTFQFSYGTKETSYGLSFFKPRPGNFRNFSVN--LYKVT 196  
 QY 495 GYY-----RKTYDNKNISNVYLDVSGYSLGVYPIDENORISFGLNADNTKLHGPRMG 549  
 Db 197 GQFPWSSURET-----DRGMSAEYSFPFWKTS-----HTVKWEG 230  
 QY 550 ISNVKQLMADGKGIOVDNNGIPDFKHDTVTYNAILGNWYSSLDPRVPFTQGMHSVDLTV 609  
 Db 231 VVRELGLSRLRTASFVRKESGHSLSKSLSHAMVI-----DSRNSILPRGALLXVQEL 285  
 QY 610 -GF--GDKTHQVVQVQNIYRPFITKKSVLGRYAKLYGNNLP-----FYENFYAGGYG 659  
 Db 286 AGYTGDSVFKEDPELQNLQIFDSVFS--ASFWGMGLVPIDGKPSIADRFLGGPT 343  
 QY 660 SVRGYDQSSSLGRSQAYLTARRGQQTTLGEVVGNGALATFGSELILPLPFK---GDWIDQ 716  
 Db 344 SIRGFSMHSIGPQSE-----GDYLGSGAYWAGGLUHLTYLPFRPQGQGFGL 390  
 QY 717 VRPVIFEGGVDFDTGMDKQTDILTQFKDQPTAEQNAKARNRPLLTQDKQLRYSAGVG 776  
 Db 391 FRTHFFLNAGNL-----CNLNYGEGPKAHRKLAEC-----IRWSYGAG 429  
 QY 777 ATW-YTPIGPLISYAKPLNKKQNDQ-TDTVQFOIGSVF 813







Best Local Similarity 19.0%; Pred. No. 0.19;		Matches 199; Conservative 125; Mismatches 319; Indels 405; Gaps 55;	
QY	9	FQVSAMTMAVMVMSTHAQAADFMANDITITGLORVTIESLQSVLPFRGLGVSENLQ---	66
DB	25	FGAASVLGLSVLGAQAQV---VKADETVAS-SEPTIAS--VAPASTEAVEAEAKTN	78
QY	67	ADGKALYATGNFSDVQVYHOGELIYQVT-ERPLIAEINFEGRNLIPK---EGLOELK	122
DB	79	AENTSAVATTS--TEVE---KAKAVLEOVTSSEPLLACL---GOKELAKTEDATLAKAIE	130
QY	123	NA--GLAVGOPL---KQATVQMITEITL---NQYISQGYNTEI-----	158
DB	131	DAQTKLAAKAILADSEATVQEAQVAQVAAKVAEALGNELQKVTVDGLLTAALDTVPD	190
QY	159	TVKOTML--DGNRVKLMT-----FAGKPARVVDINIIGNQHFSDADLDIVLAIKDNK	210
DB	191	TTASTLKVGDEGTLLDSTTTATPSMAEPNGAAIAP-----HTLRTQDG- 234	
QY	211	INPLSKADRYTOEK--LVTSLNLRK-YLNA--GFVRFEIKD-----AKLINED	256
DB	235	IKATSEPNWYTFESVDLSYKKNKMASSYKGAEDVAYIRYSLDNDSSSTAVLAEL-VSRT	293
QY	257	KNRIFVEISLHEGEQYRGQTQL----GNLTYT-----QAELEALKFKAEFGFSQAM	306
DB	294	TGDLVKYTIIEGESVTFSHPTKVNANNSNITVTYDTSLSANTPGALKFSANDDVYSTI	353
QY	307	L-----EQTNNISLTKFGD-----DGYYY-----	325
DB	354	IVPAYQINTRTYTESGKVLATYGLQTIAGQVVPSSVRFTGYDYVAATTKAVOGPYPK	413
QY	326	-----AQRVPTRIINDESRVDVEYIIDPVHPVVVRINFTGNFKTQDEVLK	372
DB	414	GTVYLACTVQKDTQYKVIKREIVENDQAVLKPYLD---PIYKGEVDWRT-----	461
QY	373	REMROLEGALASNOKIQLSRARIKMTGTFKHHVTVTRVPVNSPDQVDVNFVVEOPSGS	432
DB	462	-----DTGTGTELLT-----TSPTTYKVGTYD-----	484
QY	433	TIAAGYSOGGVTFQFDV-----SONFMGTGKHVNASFSRSETR-----EVV-	475
DB	485	-----YNINSKITADPTIDPTKNVWVFKESEQNEGSKYRVIAQNSGDETKGIYKIYI	539
QY	476	-----SLGMNPTFTYVNGVSQSLSG-----YVRK	499
DB	540	ATQVWTTKLGTNEGWEDYSDDOAGIKFNKGFAPAGVQNTLRNATPATAVETTVYKES	599
QY	500	TKY-----DNKNTSNYVLDISYGSLSYGYPIDENQRIISFGLNADNT-----	540
DB	600	SKYGDVIVEYDYGKQIYNSVVDTPKSAIGTEYNTDVRPASIIVAAAGTVYFYKEVKS	659
QY	541	---KLHGGFRMGISNVKOLMADGGKIQ-----VDNNG-----IPDFKHDTYTYNAILGWNY	588
DB	660	DSAKTGTGTVAGTTVYKYYEKAGSVNVNFVDINGKVIKAPYSDEK-----DAKPGYNY	713
QY	589	SSLDPRVPFTQGMHSHVDLTVGDKTHQKVVYQNGNIYRPFITKKSVLRG---YAKLYGN	645
DB	714	DR-----DL-----DQKLASITFEKKEY-----KLVPAGDYPVGVKVGKN	748
QY	646	NLPFTFENFYAGG-----VGSVRG-----YDOS-----SL	669
DB	749	NLIEVGNNTAKGIDPTTGKIEAGVKNKEVTVYVRAVTVGSVVVNYKDFEGWIKDPTDVS	808
QY	670	GRPSQAYLT-----ARRGQOTLT-----GEVVGGNALATFCSELILPLPKK	710
DB	809	APVGDAYTTDKPNEIITKGSRYVLPVSKTDGEENGKVGIEGTITVTYVYQKV-----	862
QY	711	GOWIDQVRPVIFIEGQV---FDTTGMKQOTIDLTQFKDQPAQEAQNAKARNPLLTQDK	767
DB	863	ANWIEIPNPVETDRPKVPYPTDPTPD-EPIDPT-----TPGTNGEVPNIP-----	908
QY	768	QLRYSAGVGATWYTPIGPLSISVAKPLN 795	

DB	909	---YVPG-----YTPVDPKONTPLKPID 928
RESULT	12	
ID	P3P_LACIC	
AC	P3P_LACLC	STANDARD; PRT; 1902 AA.
AC	P15292;	
DT	01-APR-1990 (Rel. 14, Created)	
DT	01-APR-1990 (Rel. 14, Last sequence update)	
DT	16-OCT-2001 (Rel. 40, Last annotation update)	
DE	P111-type proteinase precursor (EC 3.4.21.96) (Lactocypin) (Cell wall-	
DE	associated serine proteinase).	
GN	PRTP.	
OS	Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).	
OG	Plasmid.	
OC	Bacteria: Firmicutes; Bacillus/Clostridium group; Streptococcaceae;	
OC	Lactococcus.	
OX	NCBI_TaxID=1359;	
RN	[1]	
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 188-197.	
RC	STRAIN=SK11;	
RC	MEDLINE=893340435; PubMed=2760036;	
RA	Vos P., Simons G., Siezen R.J., de Vos W.M.;	
RT	"Primary structure and organization of the gene for a procaryotic,	
RT	cell envelope-located serine proteinase.";	
RL	J. Biol. Chem. 264:13579-13585(1989).	
CC	-!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE	
CC	GROWTH OF THE BACTERIA ON MILK.	
CC	-!- CATALYTIC ACTIVITY: Endopeptidase activity with very broad	
CC	specificity, although some subsite preference have been noted,	
CC	e.g. large hydrophobic residues in the P1 and P4 positions, and	
CC	pro in the P2 position. Best known for its action on caseins,	
CC	although it has been shown to hydrolyse hemoglobin and oxidized	
CC	insulin B-chain.	
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.	
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE	
CC	SUBTILASE FAMILY.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
EMBL	J04962; AAA03533.1; ALT_SEQ.	
PIR	A32634; A32634.	
HSSP	P00782; 2SRT.	
MEROPS	S08.019; -	
InterPro	IPR001899; Gram_pos_anchor.	
InterPro	IPR003137; PA.	
InterPro	IPR002029; Peptidase_S8.	
Pfam	PF00746; Gram_pos_anchor; 1.	
Pfam	PF02225; PA; 1.	
Pfam	PF00082; Peptidase_S8; 3.	
PRINTS	PR00723; SUBTILISIN.	
PROSITE	PS00136; SUBTILASE_ASPI; 1.	
PROSITE	PS00137; SUBTILASE_HIS; 1.	
PROSITE	PS00138; SUBTILASE_SER; 1.	
PROSITE	PS00343; GRAM_POS_ANCHORING; 1.	
KW	Hydrolase; Serine protease; Cell wall; Zymogen; Signal; Plasmid;	
KW	Transmembrane.	
SIGNAL	1 33	
PROPEP	34 187	
FT	CHAIN 188 1902	P111-TYPE PROTEINASE.
FT	DOMAIN 188 1876	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM 1877 1895	MEMBRANE ANCHOR (POTENTIAL).
FT	DOMAIN 1896 1902	CYTOPLASMIC (POTENTIAL).
FT	ACT_SITE 217 217	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE 281 281	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE 620 620	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	DOMAIN 1867 1872	CONSERVED IN GRAM-POSITIVE COCCI SURFACE



```
FT  SEQUENCE 1902 AA; 200550 MW; 87CECBAA9345F9D3 CRC64;
SQ  SEQUENCE 1902 AA; 200550 MW; 87CECBAA9345F9D3 CRC64;

Query Match 3.5%; Score 149; DB 1; Length 1902;
Best Local Similarity 19.6%; Pred. No. 0.74;
Matches 175; Conservative 116; Mismatches 356; Indels 246; Gaps 44;

Qy 4 SYFKGQVSAVTMAYMMV-MSTHAQAADFMAANDITIT---GLQVTTIESLQSVLPFRLG 58
Db 646 AYYKQKGTALTDLFKTVEMNTAQIPNDINNNVIVSPRQAGLVYDKAAIDALEKPS 705

Qy 59 QVSNQLADGKALYATGNSDVQYHQ--EGRIYQVTERPLIAEINFENRLI----- 112
Db 706 TVVAEN---GYPAVELKDTSTDKTFLTFTNRTTHTLYQ---MDSNTDINAVVTSAT 758

Qy 113 -PKEGL--OEGLNAGLAVGQPLKQATVQMIELET-----NQYISQGYNTEIIVK 161
Db 759 DPNQSVLVDKIDGAAIKAGSNITVPAGKTAQIEFTSLPKSFDQDQVVEGFLNFKGS-- 816

Qy 162 QTMLDGNRVKLDWT--FAEGKPARVVDINIIGNQHFSDADLDVLAIKDNKINPLSKADR 219
Db 817 ---DGRNLNPLMFFGWDGKIVD-SLNGITYSPAGNFGTVPPLKKN---NTGTQ 867

Qy 220 YTOEKLVSLENLRAKYLNAGFVRPEIKDAKLNINEDKNRIFVEISLH----- 267
Db 868 Y-YGGMVTDADGNKT-----VDDQATFSSDRKNALYNDISMKYVYLLRNISNVQV 915

Qy 268 -----EGQYRFQTOFLGNLTYYOAELEALLKFAEGESQOAMLEQNTNISTKFGDQ 322
Db 916 DILDQGNKVTLSSTNRKTYNAHSQQYIYYNA-PAWDGTYDYDQDGNKLT--ADGG 972

Qy 323 YYAQIRPVTRINDESRVDVEYYIDPVHPVVVRINF---TGNPKTODEVLRRERMQLE 379
Db 973 SVTYRISGVEGGDKRQVDFPFKLDKAPT-VRHVALSAKTENGKTQYLYTAEAKDDL 1031

Qy 380 GALASNQIOLSRARLMTGTFKKHVTVDTRVPNSPDQVNVFVVEQPSGSTIAAGYS 439
Db 1032 GLDAT-----KSVKTEINEVTN-----LDATF-----TDAGTTADGYT 1064

Qy 440 QSGGVTFQFVSQNNFMCTGKHV-----NASFSRSETREYVSLGTMNPFYTVNG--- 488
Db 1065 K---IETPLSDEQAQALNGDNSAELYLTDNASNATDQDASVQKPGSTFDLIVNGGGIP 1121

Qy 489 --VSQSLSGYRKTK-----YDNKNISNYVLDSYGGSLSYGYP 524
Db 1122 DKISSTTGYEANTQGGCTYTFSGTYPAADVGTYYTDAQGKKHDLNTTYDAATNFTASMP 1181

Qy 525 I-DENORISFGLNADNTKLHGRFRMGISNVKOLMADGKGIQVDNNGIPDFKH----- 575
Db 1182 VTNADYAAQVDLYAD--KAHTQLLKHFDTKVLMAPTFTDLKFNNNGSDOTSEATIKVTGT 1239

Qy 576 ---DYTVYNAILGMNYSLSLRPVPTQGMSH-SVDLTGVFGDK-----THQ 617
Db 1240 VSADTKTVN--VHTVAALD-----AQHHSVDVPVNYGDNITKVATDKDGNNTTEQ 1290

Qy 618 KVVYQGNIRP-FIKKSYLRGAKLYGNNLPFENFYAGGYVGRVYDQSSLSGPRSQAY 676
Db 1291 KTISS--YDPMLKSVT-----PDQGVKFGTNKE--NATSAKEVDPKT---GIAT 1335

Qy 677 LTARRQOTTIGEYVVGNA-----LATFGSELILPLPF-----KGDW----- 713
Db 1336 ITGKVKHPTTTLQVDGKQIPKIDDLTFSFTLDGLGLGQK-----PFGVVVGDPTQNTKTFQ 1390

Qy 714 -----IDQVRPVIFIEGGQVDFDTGMDKQITDLTQFKDPQ-----ATAEQNAK 756
Db 1391 EALSFILDAVAP-----TLSDSSTDAPVYTNPNFQITGTATDNAQ 1432

RESULT 13
YDF6_SCHPO
ID YDF6_SCHPO STANDARD; PRT; 475 AA.
AC Q10478;
```

```
YEHF_ECOLI
ID YEHF_ECOLI STANDARD; PRT; 826 AA.
AC P33341;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical outer membrane usher protein yehf precursor.
GN YEHF OR B2109.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / BHB2600;
RA Richterich P., Lakey N., Gryan G., Jaehn L., Mintz L., Robison K.,
RA Church G.M.;
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474 (1997).
[3]
RP SEQUENCE OF 307-826 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251358; PubMed=9097040;
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Kasai H., Kimura S., Kitagawa M., Kitagawa M., Makino K., Miki T.,
RA Mizobuchi K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y.,
RA Nishimoto H., Nishio Y., Oshima T., Saito N., Sempel G., Seki Y.,
RA Sivasubram S., Tagami H., Takeda J., Takemoto K., Wada C.,
RA Yamamoto Y., Horiuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392 (1996).
CC -1- FUNCTION: INVOLVED IN THE EXPORT AND ASSEMBLY OF A FIMBRIAL
SUBUNIT ACROSS THE OUTER MEMBRANE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane
(bY similarity).
CC -1- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

```
Matches 148; Conservative 92; Mismatches 306; Indels 176; Gaps 34;
QY 135 ATQMIEET-LTNOYISQGYNT-----EITVK-----QTMLDGNRVKL-----DMTFAEG 179
DB 42 ANIRLDNDQPLPGYDIDIVNKNQWGRKYELIIVKNDQETCLSRVIRKLGINSDFASG 101
QY 180 KPARVVDINIGNOHFSDADLIDVLAIKDNKINPLSKADRYTQEKLVTSLENLAKYLYNA 239
DB 102 KQCLTFEQLVQGGSYTWDIGVF-----RDLDFVPAQWVEELES 139
QY 240 GFYRFEIKDAKLINEDKNRIFVEISLHEGEYRFGOTQFLGNLTQAELEALLKFAE 299
DB 140 GYVPE-----NWERGINAFYTSYL---SQY-YSYKASGNKSTVVRNSGLNLGW 189
QY 300 EGFSQAMLEOTNN-----ISTKFGDDGYAQAIRPVRTRINDESRTVDVEYIIPVHPV 354
DB 190 QLHSDASFSTNNPNPGVWKSNTLYLERG--PAQLLGLTRVGDWYSSDI---FDSVFRG 244
QY 355 VRRINFTGNFKTQDEVLRERMRQL-----EGALASNQKILSLRAKLMRTGFKKH 403
DB 245 VR-----LFRDMQLPNSKQNFTRVQGIASQNALVTIE-----QNGF--- 282
QY 404 VTVDTRVPNSPDQV-----DYNFVVEEQPSGSSTIAAGYSQSGGVTFQFDYSON 453
DB 283 -VYQKEVPPGPFPAITDLQLAGGADLDVSKADGSGVTITLYVPAAPNN-LQPGVSKY 340
QY 454 NFMGTGKHVNASFSRSE--TREVISLGMTNPYFTVNGVSQSLSGYRRTKYDNKNISNYVL 512
DB 341 DLAAGRSHIEGASKQSDQVQAGYQYGFNN-LTLTYGGSWANNYYAFT----- 387
QY 513 DSYGGSISYGYPIDENQRIQFGLNADNTKLHGGRFMGINSVKQLMADGGKIQVNDNGLPD 572
DB 388 -----LGAGW-----NTRIG-AISVDATKSHSKDNGD-----VFQGSQYAIYN--K 427
QY 573 FKHDYTTYNALGLNYSLSLDRPVPPTQGMSSHVDLTVCFGDKTHQKVY----- 621
DB 428 FVSTSTFRFLAAWYSSRDYTRFNDHVAANKD-----NVRRENDVYDIADYQNDFG 482
QY 622 QGNIYRPFIRKSVLRGAKLGYNNLPFFENFYAGGYSVGRYDQS-SLGRPSQAY-LTA 679
DB 483 RKNFSANMSQSLPEGWGSV---SLSTLWRDYWGRSGSSKDYQLSYNNLRISYTLAA 538
QY 580 RRGQOTILGEVVGGNALATFQSEILLPLPKGDWIDQV---RPVIFIEGGQVEDTTGMDK 736
DB 539 SQAYDENHHEEKREN-----IFISIPF--DWGDDVSTPRQIYMSNSTFDDQGFAS 588
QY 737 QTIDLTQFKDQATAEQNAKAAANRPLLTQDKQLRYSAGVGATWTPIGPLSISYAKPLNK 796
DB 589 NN---TGLSGTVGSRDQFNQYGVN--LSHQHOGNETTAGANLTWNPATVNGSYQSSTY 643
QY 797 KQ 798
DB 644 RQ 645
RESULT 15
PIP_LACLC
ID PIP_LACLC STANDARD; PRT; 1902 AA.
AC P16271;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE PI-type proteinase precursor (EC 3.4.21.-) (Wall-associated serine
proteinase).
DE GN PRTP.
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OC Plasmid pW05.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WG2;
```

RX MEDLINE=88149035; PubMed=3278687;  
RA Kok J., Leenhouts K.J., Haandrikman A.J., Ledebor A.M., Venema G.;  
RT "Nucleotide sequence of the cell wall proteinase gene of  
RL Streptococcus cremoris Wg2";  
CC Appl. Environ. Microbiol. 54:231-238(1988).  
CC -1- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE  
CC GROWTH OF THE BACTERIA ON MILK.  
CC -1- CATALYTIC ACTIVITY: ENDOPEPTIDASE ACTIVITY WITH VERY BROAD  
CC SPECIFICITY, ALTHOUGH SOME SUBSITE PREFERENCE HAVE BEEN NOTED,  
CC E.G. LARGE HYDROPHOBIC RESIDUES IN THE P1 AND P4 POSITIONS, AND  
CC PRO IN THE P2 POSITION. BEST KNOWN FOR ITS ACTION ON CASEINS,  
CC ALTHOUGH IT HAS BEEN SHOWN TO HYDROLYSE HEMOGLOBIN AND OXIDIZED  
CC INSULIN B-CHAIN.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE  
CC SUBTILASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M24767; AAA17677.1; -  
DR HSP; P00782; I501.  
DR MEROPS; S08.019; -  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR003137; PA.  
DR InterPro; IPR000209; Peptidase\_S8.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF02225; PA; 1.  
DR Pfam; PF00082; Peptidase\_S8; 3.  
DR PRINTS; PR00723; SUBTILISIN.  
DR PROSITE; PS00136; SUBTILASE ASP; 1.  
DR PROSITE; PS00137; SUBTILASE HIS; 1.  
DR PROSITE; PS00138; SUBTILASE SER; 1.  
DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; 1.  
DR Hydrolase; Serine protease; Cell wall; Zymogen; Signal; Plasmid;  
KW Transmembrane.  
FT SIGNAL 1 33  
FT PROPEP 34 187  
FT CHAIN 188 1902 PI-TYPE PROTEINASE.  
FT DOMAIN 188 1876 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 1877 1895 MEMBRANE ANCHOR (POTENTIAL).  
FT DOMAIN 1896 1902 CYTOPLASMIC (POTENTIAL).  
FT ACT\_SITE 217 217 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 281 281 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 620 620 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT DOMAIN 1867 1872 CONSERVED IN GRAM-POSITIVE COCCI SURFACE  
SQ SEQUENCE 1902 AA; 199910 MW; 2901C7F19B2E5D0B CRC64;

Query Match 3.5%; Score 146; DB 1; Length 1902;  
Best Local Similarity 19.8%; Pred. No. 1.1;  
Matches 176; Conservative 116; Mismatches 354; Indels 244; Gaps 45;  
QY 4 SYFKGQVSANTMAYMMV-MSTHAQAADFANDITIT-----GLQRTVIESLQSVLPRLG 58  
DB 646 AYYKQLKGTALTDFLKVTVMNTAQPINDINNVIVSPRRQAGLVVDVKAIDAILEKNPS 705  
QY 59 QVWSENOL-ADGVKALVATG-----NFSDVQVYHQEGRIIYQVTERPLIAEINFEGNRLI 112  
DB 706 TVVAENGYPAVELKDFSTDKTLFTNSTTH-----ELTYQM-----DSNTDTNAVY 754  
QY 113 -----PREGI--QEGLNAGLAVGQPLKQATVQMIEITLT-----NQVISQGYNTE 157  
DB 755 TSATDPNSGVLYDKIDGAAIKAGSNITVPAGKTAQIEFTLSLPKSPDQQQFVSGFLNFK 814  
QY 158 ITVKQTMLDGNRVKLDMTFAGGKPARVVVDINIIGNQHFSDADLIDVLAIDKNKINPLSKA 217

DB 815 GS-----DGSRLNLPY-----MGFFGD--WMDGKIVDSL-----NGITYSPAG 850  
QY 218 DRYTOEKLVTSLENLRAKYLNAGFV-----RFEIKDAKLNTINEDKNRIF----- 261  
DB 851 GNFGTVPLLTN-KWTGTOYY--GGWVTADAGNQTVDDQAIASFSSDKNALINDISMKYLLR 908  
QY 262 ----VEISLHEGEQYRFGOTQFLGNL--TYTQAELEALLKFAEFGFSQAMLEQTNNTIS 315  
DB 909 NISNVQVILDGQGNKVTTLSSTNLTNTYNAHSQOYIYNA--PAWDTGYVDQDQGNIK 967  
QY 316 TKFGDDGYYIAQIRPVTRINDESRTVDVEYIDPVHPVYVVRINF---TGPKTQDEVLR 372  
DB 968 T--ADDSYTYRISGVPPEGDKQVDFPFKLDKAPT--VRHVALSAKTENGKTQYIYLA 1024  
QY 373 REMRLEGALASNOKIQLSRARLMRTGFFKHVTVTRPVNPSDQVDVNFVVEEQPSGSS 432  
DB 1025 EAKDLSG-----LDATKSVKTAI-----NEVTNLDATF-----TDAG 1057  
QY 433 TIAAGYSQSGGVTFQFVDSQNNFMGTGKHV-----NASFSRSETREYISLGMTNPYFT 485  
DB 1058 TTADGYTK---IETPLSDEQAQALGNGDNSAELYLTDNASNATDQDASVQKPGSTFSDLI 1114  
QY 486 VNG-----VSQSLSGYYRKT-----YDN-----KNISNYVLDVSGG 517  
DB 1115 VNGGIPDKISSTTGYEANTQGGGYTFTSGTYPAADVGTYTNAQGGKHDLTNTTYDAATN 1174  
QY 518 SLSYGYPI-DENQRISFGINADNTKLHGGRFMGISNVKOLMADGGKIOVDNNGIPDFKH- 575  
DB 1175 SFTASMPVTNADYAAQVDLYAD--KAHTQLLKHFDTKVLRTAPTFTDLKFNNGSDOTSEA 1232  
QY 576 -----DYTTYNAILGNWYSSLDPRVPFTQGMSH-SVDLTVGFGDK----- 614  
DB 1233 TIKVTGTVSADTKTVN--VGDVVAALD-----AQHHFSVDVPVNYGONTIKVTATDED 1283  
QY 615 ----THQKVYVQGNLYRP-FIKKSVLRGYAKLGYGNLPPFYENFVAGGYGVRGVDQSSL 669  
DB 1284 GNTTEQKTISS--YDPMLKNSVT-----FDQGVTFGANEF--NATSAKFYDPKT- 1331  
QY 670 GPRSQAYLTARRGOQTTLGEVVGNA-----LATFGSELILPLPKGDWIDO 716  
DB 1332 --GIATITGVKVKHPTTTLQVDGKQIPIKDDLTFFSFTLDTLGLGK-----PF----- 1376  
QY 717 VRPFIIEGQVFDTTGMDKQOTIDLTQFKDQATAEQNAKAAANRPLLTQD 766  
DB 1377 -----GVVVGDTQNKTFQEAFTILDVAVAPTLLESSTADAPVYIND 1418

Search completed: September 5, 2002, 10:03:36  
Job time: 529 sec

---





```
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE SURFACE ANTIGEN.
GN VC2252.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA McElane M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483(2000).
DR EMBL: AE004297; AAF95396.1; -.
DR TIGR: VC2252;
DR InterPro; IPR000184; Bac_surfAg_P15.
DR Pfam; PF01103; Bac_surface_Ag; 1.
KW Complete proteome.
SQ SEQUENCE 803 AA; 89930 MW; 1998B3838F6C041D CRC64;

Query Match 28.6%; Score 1203; DB 16; Length 803;
Best Local Similarity 33.4%; Pred. No. 4.5e-58;
Matches 278; Conservative 155; Mismatches 328; Indels 72; Gaps 15;

QY 16 MAVVMVMTAAQAD-FMADITITGLQVVTIESLQSVLPFRLGQVSVENQADGVKALY 74
DB 8 LATLTSVANGAEKFFVQDIDQLQRLVGAALLKMPVRGSDVSDQVANIKA 67
QY 75 ATGNFSDVQVHQEIRIIVQTERPIAEINFEGRNLPKEGLQELKNAGLAVGQPLQ 134
DB 68 SSGNFEDVKVLRDGNLTAVQVKERTIASVSFSGNKAKEQLKQNLKLEASSIRVGEALDR 127
QY 135 ATVQMIETELNQYISQGYNTIEIVKQTMDCNRVLDKMTFAEGKPARVVDINIIGNOH 194
DB 128 TTLSNIEKLEDFYISVGNATKAVVTPLPNRNADLKFEVTEGSAKIQINFIQNOV 187
QY 195 FSDADLDVATKDNKI--NPLSKADRYTOEKLVTSLNLRKAKYLNAGVREIKDAKLN 252
DB 188 FSDELLSRFNLDVAVNWFNAD-DKYKKQVLADIEALRYYLDRGLKFEQVDSQVA 246
QY 253 INEDKNRIFVEISLHEGEQYRFGQTQFLGNLTYYTQAEALKFKAEQFSQAMLEQTIN 312
DB 247 ISPDKKGVITLNLNEGEPTYVSKVQFRGELMGKEAFTSLIPFEIGETNGSVATRLKE 306
QY 313 NISTKFGDDGYAIIQRPVTRINDSRVVDVYIIPVHPVYVRRINFTGNFKTQDEVLR 372
DB 307 SVKKVLGSEGYAYPQVRIPEFDDEKQVSLVHVHVEAGKRVYVRDIREVGNNSRDEVLR 366
QY 373 REMROLEGALASNOKIQLSRARLMTGFKHVTVDTPVPNSPDQVDVNFVVEEQPSGSS 432
DB 367 REMRQWEGSWLNKDIETGKTLNRLRGFFETVEVQTVRVPGEQDQDLVYSVKEANSNV 426
QY 433 TIAAGYSGGVTFQFDVSNQNNFMGTGKHVNASFSRSETREYVSLGNTNPTFTVNGVSQS 492
DB 427 NFGVGYGTESGVYFQVGLQDQNFLLSGNRVGNVAMINDYQKNLTLEYRDPYWNLDGVSIG 486
QY 493 LSGYRKYKYNKNTSNVLDVSGSLSYGYPIDENQRIISFGLNADNTKLHGRFWMGTSN 552
DB 487 GKVFYVQFEASRAGIVDYTNESYGTSLWGGFPFDELNRFEGIGYTHNKI--GNLTPIYQ 544
QY 553 VKQLMA-----DGGKIQVONNGIPDFKHDYTTYNAILGNWYSSLDPRVFPFTQGMHS- 604
DB 545 VENFLAAQASNDISGNNLLTDD-----FDINLSWTRNRLNNSYFPTAG-NHQR 591
```

---

```
QY 605 --VDLTVGFCDKTHQKVYQGNIRPFIKKS-----VLRGYAKLGYGNN-----LP 648
DB 592 AFYKMTVPGSDAQYFKLQYDVRYQYFPLTKKHFTLLRG--RLGYNGYQOTDGKONLFP 649
QY 649 FYENFYAGYGSVGYDQSSSLGPRS--QAYLTARRGOQTTLGEVVGGNALATFGESELILP 706
DB 650 FYENFYAGGFTSLRGFGSNSAGPKAVYRDYSGSNNGSDTATDDSDVSGNATLASVELIIVP 709
QY 707 LPFKGDWI--DOVRPVLIFEGGOVEDTGMKDQID-----LTOFKDQPOATAEQNAKANR 760
DB 710 TPFASEARNQRTSIFYDMASVMDTEFDYRGKADYGNQYVYDSDP----- 756
QY 761 PLITDQKRLYSAGVGATWYTPIGPLSISYAKPLNKKQNDQTDVTVQFQIGSVF 813
DB 757 -----TNYRSSVGVALQWVSPMGPLVSLAKPIKKYEGDDEEFFFTIGRTF 803

RESULT 4
Q9R2E3 PRELIMINARY; PRT; 785 AA.
AC Q9R2E3;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE HYPOTHETICAL 87.9 KDA PROTEIN.
GN YZZN.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA Takenoto K., Mori H., Murayama N., Kataoka K., Yano M., Itoh T.,
RA Yamamoto Y., Inokuchi H., Miki T., Hatada E., Fukuda R., Ichihara S.,
RA Mizuno T., Makino K., Nakata A., Yura T., Sampei G., Mizobuchi K.;
RT "Systematic sequencing of the Escherichia coli genome: Analysis of the
RT 4.0 - 6.0 min (189,987 - 281,416bp) region.";
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=94261430; PubMed=8202364;
RA Fujita N., Mori H., Yura T., Ishihama A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of the
RT 2.4-4.1 min (110,917-193,643 bp) region.";
RL Nucleic Acids Res. 22:1637-1639(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=82059454; PubMed=6272196;
RA An G., Bendak L., Mamelak A., Friesen J.;
RT "Organization and nucleotide sequence of a new ribosomal operon in
RT Escherichia coli containing the genes for ribosomal protein S2 and
RT elongation factor Ts.";
RL Nucleic Acids Res. 9:4163-4172(1981).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=93077430; PubMed=1447125;
RA Yamanaka K., Ogura T., Niki H., Hiraga S.;
RT "Identification and characterization of the smbA gene, a suppressor of
RT the mukB null mutant of Escherichia coli.";
RL J. Bacteriol. 174:7517-7526(1992).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA Smallshaw J., Kelln R.;
RT "Cloning, nucleotide sequence and expression of the Escherichia coli
RT K-12 pyrH gene encoding UMP kinase.";
RL Genetics 11:59-65(1992).
RN [6]
```

RX SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=94240115; PubMed=8183897;  
 RA Janosi L., Shimizu I., Kaji A.;  
 RT "Ribosome recycling factor (ribosome releasing factor) is essential  
 for bacterial growth.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:4249-4253(1994).  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=91317739; PubMed=1860827;  
 RA Shimizu I., Kaji A.;  
 RT "Identification of the promoter region of the ribosome-releasing  
 factor cistron (fir)." ;  
 RL J. Bacteriol. 173:5181-5187(1991).  
 RN [8]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=90062117; PubMed=2684966;  
 RA Ichikawa S., Kaji A.;  
 RT "Molecular cloning and expression of the ribosome releasing factor." ;  
 RL J. Biol. Chem. 264:20054-20059(1989).  
 RN [9]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=86008268; PubMed=2995358;  
 RA Icho T., Sparrow C.P., Raetz C.R.H.;  
 RT "Molecular cloning and sequencing of the gene for CDP-diglyceride  
 synthetase of Escherichia coli." ;  
 RL J. Biol. Chem. 260:12078-12083(1985).  
 RN [10]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=91123198; PubMed=1991717;  
 RA Hirvas L., Koski P., Vaara M.;  
 RT "The ompH gene of yersinia enterocolitica: cloning, sequencing,  
 expression, and comparison with known enterobacterial ompH  
 sequences." ;  
 RL J. Bacteriol. 173:1223-1229(1991).  
 RN [11]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=91100302; PubMed=1987124;  
 RA Dicker I., Seetharam S.;  
 RT "Cloning and nucleotide sequence of the firA gene and the firA200(ts)  
 allele from Escherichia coli." ;  
 RL J. Bacteriol. 173:334-344(1991).  
 RN [12]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=88058790; PubMed=2824445;  
 RA Crowell D., Reznikoff W., Raetz C.;  
 RT "Nucleotide sequence of the Escherichia coli gene for lipid A  
 disaccharide synthase." ;  
 RL J. Bacteriol. 169:5727-5734(1987).  
 RN [13]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=88139188; PubMed=3277952;  
 RA Coleman J., Raetz C.;  
 RT "First committed step of lipid A biosynthesis in Escherichia coli:  
 sequence of the lpxA gene." ;  
 RL J. Bacteriol. 170:1268-1274(1988).  
 RN [15]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;

RX MEDLINE=92380982; PubMed=1355089;  
 RA Li S., Cronan J.;  
 RT "The genes encoding the two carboxyltransferase subunits of  
 Escherichia coli acetyl-CoA carboxylase." ;  
 RL J. Biol. Chem. 267:16841-16847(1992).  
 RN [16]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=90094229; PubMed=1688424;  
 RA Zhou Z., Syvanen M.;  
 RT "Identification and sequence of the drpA gene from Escherichia coli." ;  
 RL J. Bacteriol. 172:281-286(1990).  
 RN [17]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=90370122; PubMed=2203971;  
 RA Eriani G., Delarue M., Poch O., Gangloff J., Moras D.;  
 RT "Partition of tRNA synthetases into two classes based on mutually  
 exclusive sets of sequence motifs." ;  
 RL Nature 347:203-206(1990).  
 RN [18]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=93094132; PubMed=1459951;  
 RA Gervais F.G., Drapeau G.;  
 RT "Identification, cloning, and characterization of rcsF, a new  
 regulator gene for exopolysaccharide synthesis that suppresses the  
 division mutation ftsZ84in Escherichia coli K-12." ;  
 RL J. Bacteriol. 174:8016-8022(1992).  
 RN [20]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=94124004; PubMed=7904973;  
 RA Allikmets R., Gerrard B., Court D., Dean M.;  
 RT "Cloning and organization of the abc and mdl genes of Escherichia  
 coli: relationship to eukaryotic multidrug resistance." ;  
 RL Gene 136:231-236(1993).  
 RN [21]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=93011013; PubMed=1396599;  
 RA Condon C., Phillips J., Fu Z., Squires C., Squires C.;  
 RT "Comparison of the expression of the seven ribosomal RNA operons in  
 Escherichia coli." ;  
 RL EMBO J. 11:4175-4185(1992).  
 RN [22]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RA Nakayashiki T., Adachi T., Miki T., Inokuchi H.;  
 RT "Nucleotide sequence of the rnhH gene from Escherichia coli." ;  
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
 RN [23]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=79012454; PubMed=358189;  
 RA Young R.A., Steitz J.A.;  
 RT "Complementary sequences 1700 nucleotides apart form a ribonuclease  
 III cleavage site in Escherichia coli ribosomal precursor RNA." ;  
 RL Proc. Natl. Acad. Sci. U.S.A. 75:3593-3597(1978).  
 RN [24]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=93116053; PubMed=1474579;  
 RA Cormack R., Mackie G.;  
 RT "Structural requirements for the processing of Escherichia coli 5 S  
 ribosomal RNA by RNase E in vitro." ;



OC	Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OX	Xylella.
OC	NCBI_TaxID=2371;
RN	[1]
RN	SEQUENCE FROM N A.
RP	SPRAIN-9A5C.
RC	MEDLINE=20365717; PubMed=10910347;
RX	Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA	Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA	Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA	Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carreir H.,
RA	Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA	Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA	Faciniani A.P., Franca S.C., Franco M.C., Ferreira V.C.A., Ferro J.A.,
RA	Fraga J.S., Frasca A.A., Goldman M.H.S., Gomes S.L., Gruber A.P.,
RA	Garnier M., Goldsm G.H., Goldman M.H.S., Kempner E.L., Kitajima J.P.,
RA	Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper S.L., Kitajima J.P.,
RA	Krieger J.E., Kuranee E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA	Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA	Macado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA	Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA	Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA	Moore D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA	Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA	de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA	Paixoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesquero J.B.,
RA	Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA	de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA	da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA	da Silveira J.F., Silverstein M.Z., Siqueira W.J., de Souza A.A.,
RA	de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
RA	Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA	Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT	"The genome sequence of the plant pathogen Xylella fastidiosa.";
RL	Nature 406:151-159(2000)
DR	EMBL; AE003941; AAF83856.1; -
DR	InterPro; IPR000184; Bac_surfAg_D15.
DR	Fram; PF01103; Bac_surfaceAg; 1.
KW	Complete proteome.
SQ	SEQUENCE 784 AA; 87991 MW; 719331DD8E18192 CRC64;

  

Query Match	27.9%; Score 1172.5; DB 16; Length 784;
Best Local Similarity	33.8%; Pred. No. 2.1e-56;
Matches	278; Conservative 151; Mismatches 321; Indels 73; Gaps
QY	14 MTMVMVMTSHAQAAD-FMANDITIGLQRVTIESLSQVLPRFGOVSVENQLADGVKA 72
Db	: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db	12 LTIAANFLPVLTAQSFANDIRVDGLRIASGTVFTPLVNRGTVDVDAKVVADIRA 71
QY	73 LYATGNFSDVOVYHOEGRIIVQVTERPLIAEFNEGRNLIPKEGLQGKNAGLAVGQPL 130
Db	: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db	72 LYRTGFENVRIDRQNILVVKVERAINKLITIGNKDIKSELLKGLSEIGSGGTF 131
QY	133 KOATVOMIETELTNOYISOGYNTTEITVKTMLDGNRVKLDMTFAEGKPARYVDINIIGN 190
Db	: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db	132 DRSLDRVTOELKRQNNRGKYNVQMTTTTPLDRNRVDVTVAIKEGKAAKIRHINLIGT 191
QY	193 QHFSDDLIDVIAIKD-NKTNPLSKADRYOQEKLVTSLBNLRKAYLNAGVFREIKDAKL 251
Db	: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db	192 EKFNKNDVMSAWESKEHNASWYRRDQYSKEKLSGDKLEKINSWLYDRGVDFDNIDSTQV 251
QY	252 NINEDKNRFVEISLHGEQYRFQOTQFLGNLTQTQAELEALKFKAEEGFSQAMLBOTT 311
Db	: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db	252 SISPEKHMMFITAGVTGGDQTKISSIKVTGNTVLDPQBIEKLVLPKTGDIISRVLLEYSS 311
QY	312 NNISTKFGDGGYYAAQIRPVTRINDERTSDVEYIIDVPVHPVYVRINFTNGFKTODEVL 371
Db	: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db	312 AAINTLSNIGYAFSKVNPPTANRADRTVAVNLHVPGPVTVRQLIFKGNFTSDEVL 371
QY	372 REMRQLEGALASNQIKQLSRALMRMGTFPKHTVTDTRFPVNSPDQVDNVFFVEEPGSG 431
Db	: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db	372 REMROFENSWSOAADRSKRIORUGYEADVDESTPPGNSDQVDVIDVTTWKETTSGS 431



RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,  
RA Whitehead S., Spratt B.G., Barrell B.G.;  
RT "Complete DNA sequence of a serogroup A strain of *Neisseria*  
RT meningitidis 22491";  
RL Nature 404:502-506(2000).  
DR EMBL; AL162752; CAB83401.1; --  
DR InterPro; IPR000184; Bac\_surfAg\_D15.  
DR Pfam; PF01103; Bac\_surface\_Ag; 1.  
KW Complete proteome.  
SQ SEQUENCE 797 AA; 88404 MW; 65DE47E00C9E1D1F CRC64;

Query Match 26.4%; Score 1111; DB 16; Length 797;  
Best Local Similarity 32.4%; Pred. No. 5.1e-53;  
Matches 265; Conservative 155; Mismatches 361; Indels 38; Gaps 14;

QY 10 QVSAMTAMVMYSTHAQAADFMANDIITGLQRTVIESLSVLPRLGOVYSENOLADG 69  
DB 2 KLKQIASALMLGISPLALADFTIQDIRVEGLQRTPESTVFVLPVKVGDYNDTHGSAI 61  
QY 70 VKALYATGNFSDVQVYHQBGRRIYQVTERPLIAEINFEGNRLIPKEGLQEGKLNAGLAVG 129  
DB 62 IKSLYATGDFDVRVETADGQLLTAVIERPTIGSLNITGAKMLQNDAIKKNLESFGLAQS 121  
QY 130 QPLKQATVOMIETELTNOYISOGYYNTEITVKQTMLDGNRVKLDMTFAEGKPARVVDINI 189  
DB 122 QYFNQATNOAVAGLKEEYLGKGLNIQITPKVTKLARNRVDIDITIDEGKSAKITDIEF 181  
QY 190 IGNOHFSADLDIVLAIKDKNI-NPLSKADRYTQEKLVTSLENLRAKYLNAGVFEIKD 248  
DB 182 EGNQVYSORLKMRLMSTEGGLTWTLSNQFNQEKFAQDMKVTDFYQNNGYDFRILD 241  
QY 249 AKLNEDKNRIFVVISLHEGEQYRFGQTQFLGNLT-YTQAELEALLKFAEGFSQAML 307  
DB 242 TDQTNEDKTQTKITITVHEGGRFVWGVSEIGDNEVPKAELEKLLTMKPKGYERQOM 301  
QY 308 EOTTNNISTKFGDDGYIAQIRPVTRINDESRVTVDEYVIDPVHVPVYRRINFTGNFTQ 367  
DB 302 TAVLGEIQNRMGSAAYSEISVQPLPNAETKTVDFVHLIEFGKRIYVNEIHTGNKTR 361  
QY 368 DEVLRREROLEGALASNOKIOLSRARLMRTGFFKHVTVTRPVNPSDDVDVNFVEEQ 427  
DB 362 DEVVRELRLQMSAPYDTSKLQSKRVERELLGYFDNVQFDVAVPLAGTDPKVDLNNLSLTER 421  
QY 428 PSGSSTIAAGYSQSGVTFQFDVSONNEMFGTKHYNASFSRSETRVYSLGNTNPFYTVN 487  
DB 422 STGSILDLSAGVQDFTGLVMSAGVSQDNLFGTGKSAALRASRSKTTLNGSLSFDPYFTAD 481  
QY 488 GVSQSLSGYR--KTKYDNKNISNVLDYSQGSLSYGPIDENQRIISFGLNADNTKLHGG 545  
DB 482 GVSGLDYVYKAFDPRKASTSIKQYKTTTAGAGIRMSVPVTEYDRVNEGLVAEHLTVN-- 539  
QY 546 RFMGISNVKQLMADGGKIQVDNNGIPDFKHDTYTTNAILGNYSLSLDRPVPTQGMHSV 605  
DB 540 ---TYNKAPKHVADPIKKGKTDG--TDGSFQGLWYKGTGVGWRNKTDLSALMTRGYLTGV 595  
QY 606 DLTGVF-GDK--THQKVYVQGNIRYPIKKSVLK----GYAKLGYG--NNLPFFYENFYAG 656  
DB 596 NAEIALPGSKLQYYSATHNQWTFPPLSKTFTLMLGGEVGIAG-GYGRTEKEIPFFENFYGG 654  
QY 657 GYGSVRGVDQSSLSGPRSQAYLTARGQQTTLGEVY--GGNALTATGSELILPLPKPGDWI 714  
DB 655 GLGSRVYGESGTGLPKP-----VYDEYGEKISYGNKKANYSABELFPMPGAKD-A 703  
QY 715 DQVRVIEEGQVDFDTGMDKQITDLTQFDKPDQATQNAKANAANRPLLTQDKQLRYSG 774  
DB 704 RTVRLSLFADAGSVWVGKTYDDNSSAT-----GGRVQNIYAGNTHKSTFTNELRYSG 758  
QY 775 VGATWYTPIGPLSLSYAKPLANKQNDQTDTVQFQIGSVF 813  
DB 759 GAVTWLSPGLPKMSYATPLAKKKPEDEIQRFQGLGTTF 797

RESULT 8  
O30912 PRELIMINARY; PRT; 797 AA.  
AC O30912;  
DT 01-JAN-1998 (TremBLrel. 05, Created)  
DT 01-JAN-1998 (TremBLrel. 05, Last sequence update)  
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)  
DE OUTER MEMBRANE PROTEIN OMP85.  
GN OMP85.  
OS *Neisseria meningitidis*.  
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; *Neisseria*.  
OX NCBI\_TaxID=487;  
RN [1]  
RC SEQUENCE FROM N.A.  
RC STRAIN=HH;  
RA Manning D.S., Reschke D.K., Judd R.C.;  
RT "Omp85 proteins of *Neisseria gonorrhoeae* and *Neisseria meningitidis*  
RT are similar to *Haemophilus influenzae* D-15-Ag and *Pasteurella*  
RT *multocida* Omp87";  
RL Microb. Pathog. 23:0-0(1998).  
DR EMBL; AF021245; AAC17599.1; --  
DR InterPro; IPR000184; Bac\_surfAg\_D15.  
DR Pfam; PF01103; Bac\_surface\_Ag; 1.  
SQ SEQUENCE 797 AA; 88539 MW; CF911B5F70B999CF CRC64;

Query Match 26.4%; Score 1109; DB 2; Length 797;  
Best Local Similarity 32.4%; Pred. No. 6.6e-53;  
Matches 265; Conservative 154; Mismatches 362; Indels 38; Gaps 14;

QY 10 QVSAMTAMVMYSTHAQAADFMANDIITGLQRTVIESLSVLPRLGOVYSENOLADG 69  
DB 2 KLKQIASALMLGISPLAFADFTIQDIRVEGLQRTPESTVFVLPVKVGDYNDTHGSAI 61  
QY 70 VKALYATGNFSDVQVYHQBGRRIYQVTERPLIAEINFEGNRLIPKEGLQEGKLNAGLAVG 129  
DB 62 IKSLYATGDFDVRVETADGQLLTAVIERPTIGSLNITGAKMLQNDAIKKNLESFGLAQS 121  
QY 130 QPLKQATVOMIETELTNOYISOGYYNTEITVKQTMLDGNRVKLDMTFAEGKPARVVDINI 189  
DB 122 QYFNQATNOAVAGLKEEYLGKGLNIQITPKVTKLARNRVDIDITIDEGKSAKITDIEF 181  
QY 190 IGNOHFSADLDIVLAIKDKNI-NPLSKADRYTQEKLVTSLENLRAKYLNAGVFEIKD 248  
DB 182 EGNQVYSORLKMRLMSTEGGLTWTLSNQFNQEKFAQDMKVTDFYQNNGYDFRILD 241  
QY 249 AKLNEDKNRIFVVISLHEGEQYRFGQTQFLGNLT-YTQAELEALLKFAEGFSQAML 307  
DB 242 TDQTNEDKTQTKITITVHEGGRFVWGVSEIGDNEVPKAELEKLLTMKPKGYERQOM 301  
QY 308 EOTTNNISTKFGDDGYIAQIRPVTRINDESRVTVDEYVIDPVHVPVYRRINFTGNFTQ 367  
DB 302 TAVLGEIQNRMGSAAYSEISVQPLPNAETKTVDFVHLIEFGKRIYVNEIHTGNKTR 361  
QY 368 DEVLRREROLEGALASNOKIOLSRARLMRTGFFKHVTVTRPVNPSDDVDVNFVEEQ 427  
DB 362 DEVVRELRLQMSAPYDTSKLQSKRVERELLGYFDNVQFDVAVPLAGTDPKVDLNNLSLTER 421  
QY 428 PSGSSTIAAGYSQSGVTFQFDVSONNEMFGTKHYNASFSRSETRVYSLGNTNPFYTVN 487  
DB 422 STGSILDLSAGVQDFTGLVMSAGVSQDNLFGTGKSAALRASRSKTTLNGSLSFDPYFTAD 481  
QY 488 GVSQSLSGYR--KTKYDNKNISNVLDYSQGSLSYGPIDENQRIISFGLNADNTKLHGG 545  
DB 482 GVSGLDYVYKAFDPRKASTSIKQYKTTTAGAGIRMSVPVTEYDRVNEGLVAEHLTVN-- 539  
QY 546 RFMGISNVKQLMADGGKIQVDNNGIPDFKHDTYTTNAILGNYSLSLDRPVPTQGMHSV 605  
DB 540 ---TYNKAPKHVADPIKKGKTDG--TDGSFQGLWYKGTGVGWRNKTDLSALMTRGYLTGV 595  
QY 606 DLTGVF-GDK--THQKVYVQGNIRYPIKKSVLK----GYAKLGYG--NNLPFFYENFYAG 656  
DB 596 NAEIALPGSKLQYYSATHNQWTFPPLSKTFTLMLGGEVGIAG-GYGRTEKEIPFFENFYGG 654





RT	"Complete genomic sequence of <i>Pasteurella multocida</i> Pm70.";
RL	PROC. Natl. Acad. Sci., U.S.A. 98:3460-3465 (2001).
DR	EMBL; AE006235; AAK04076.1; -
DR	InterPro: IPR000184; Bac_surfAg_pl5.
DR	InterPro: IPR001064; Crystallin.
DR	Pfam: PF01013; Bac_surface_Ag; 1.
DR	PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
DR	Hypothetical protein; Complete proteome.
SW	SEQUENCE 791 AA; 87761 MW; 024550DD8F99CD82 CRC64;

Query Match	25.0%	Score 1049;	DB 16;	Length 791;
Best Local Similarity	31.3%;	Pred. No. 1.3e-49;		
Matches 258;	Conservative 162;	Mismatches 344;	Indels 60;	Gaps 17;

14	QY	MTNAVMMVNSTHQAADFMANDJITITGLQRTVITESQSVLPFRLGQVVSVENQJADGVKAL	73
15	QY	LLASLFGSTTAAFAFPVVKDIPRVDGVOAGTSGSVLATIPRVGQARATDNDIANVVRKL	63
74	QY	YATGNFSDVOVYHQEGRIIYQVTERPLLIAINFEGNRLIPKEGLQBLKNAGLAVGQPLK	133
64	Db	FLSQYDDVKASREGNLTLYVVMKPVISNVVDIGNKSPIDEAIKQNLDAENGFKVGSDVLN	123
134	QY	QATVQMIETELTNOYISQGYTEITITVQOTMLDGNVKLDMTFAEGKPARVVDINTIGNQ	193
124	Db	RAKLEEFRAKGIYVHNSVGRYNAKVEALVNTLPNNSAEIKIOINEDDVALFKEIIFEQNO	183
194	QY	HFSDDLIDVLAJDKNKINPLSKADRYTOEQLVTSLENLRAKYLNAGFVRFETKDAKLM	253
184	Db	AFSSKLEQOMELQDAWKLP-GNFKDQTOFNKDLIETLRSYLDYRGYAQOFLDPIK	242
254	QY	NEDKNRIFVEISLHEGEQYRFQTOPLGNLITYQAELEALLFKABEG-FSOAMLBOITTN	312
243	Db	SDDKKEARVLIKVEGLDYTVKSARILGDVGMSAELAPILDAIQNLGIFRANVLVEQ	302
313	QY	NISTKFGDGYAAQTPRTRINDESRITVDVEYIDPVHPVYVRRINFTGNFKTODEVL	372
303	Db	RKSKLGEKGYATAQVNVHPTDEODKTLSDLDFIVEAGKSYTVRQIRFEGNTSSADSTLR	362
373	QY	REMRQLEGALASNOKIQLSRARLMRTGFKFHVTVDFRPVNSPDQVDVNFVVEEOPSGS	432
363	Db	QEMRQEGAWLSSELVELGKCLDRDGFESFVETKTEAIPGS-DQVDVYIKVKERTGT	421
433	QY	TIAAGYSQGGVTFQPDVSONPMFGKHKVNASFSRSETREYISLGMTNPYTFVANGVSOS	492
422	Db	NFGIGYTESGLSYQASIKQDNFLGMSGSISLGGTRNDYGTITNLGYNPEYFTKDGVS-S	479
493	QY	LSGIYRKTYDKNKNISN--YVLDSYGGSLSYGYPIDENQRIISFGLNADNTKLHGFRMG	549
480	Db	LGNVFEIDSKSNTSAAYGRTSYGNLTIGFPVNNENSYLGVGYTNKL-----	532
550	QY	ISNV-KQLMADGGKIQVDNNGIPDEF-HDYTYTNALIGWNYSLDRFPVPTQGMSSHVD-	606
533	Db	-KNIAPEYNRDLYRQSMKYNDSWTKSHD--FDLSFGHNYNSLNRGYPFTKCVRANTGG	588
607	QY	-LTVFGDKTHOKVYVQGNIRPF-----IKKSVLRGYAKLGYGNLNPYENFY	654
589	Db	RTVITPSDNKKYIKLNAEAGFYPLDREHGWLSSRISASFADGFG----GKRLPFQYVS	644
655	QY	AGGYSVRGYQDSSILGPRSOAYLTARRGOOTTL--GEVVGGNALATFGSELILPFPKGD	712
645	Db	AGGILSLRGFAGAIQPN-ITYRTQCPDYSCLVSDSDVIGGNAMWTASTELIVTPFPFVAD	703
713	QY	W-IDOVRPVIFIEGGGVDTTGMKDQITDLTOFKDQPOATEQNAKAA--NRPLLTQDKQL	769
704	Db	KNQNSVRTSLFVDAASVWNT-----RWKAEDKAKFAXINLPDYSDPSRV	747
770	QY	RYSAGVGATWKTPIGPLSTSYAKPLNKKQNDQTDIVQIQISVF	813
748	Db	RASAGVALOMQSPGLVFWFSYAKPLUKKYQOGEIQFSGIGTF	791

RESULT 13

```

Q51930
ID Q51930 PRELIMINARY; PRT; 789 AA.
AC Q51930; Q51921;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE OUTER MEMBRANE ANTIGEN OMA87.
GN OMA87.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OC NCBI_TaxID=747;
RN [1]
RS SEQUENCE FROM N.A.
RC STRAIN=PEA100;
RA Ruffolo G.C., Adler B.;
RL Infect. Immun. 64:0-0(0).
RN [2]
RS SEQUENCE OF 750-789 FROM N.A.
RC STRAIN=9222;
RC MEDLINE=95369730; PubMed=7642134;
RA Delamarche C., Manoha F., Behar G., Houlgatte R., Hellman U.,
RA Wroblewski H.;
RT "characterization of the Pasteurella multocida skp and fira genes.";
RL Gene 161:39-43(1995).
DR EMBL; U60439; AAC44600.1; -.
DR EMBL; X74357; CAA52399.1; -.
DR InterPro; IPR000184; Bac_surfAg_D15.
DR InterPro; IPR001064; Crystallin.
DR Pfam; PF01103; Bac_surfaceAg; 1.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 789 AA; 87437 MW; DC1D004AAFB2C09E CRC64;

```

Query Match 23.8%; Score 1000; DB 2; Length 789;  
Best Local Similarity 30.8%; Pred. No. 6.4e-47;

QY	14	MTAVMVMSTHAAQADFMANDITITGLQRTVITSLQSIVLPRFLQGVVSENQLADGVKAL	73
Db	4	LLIASLLFGSTTAAAFVVKDIRVDGVQAGTEGSSVLATLPRVGQATRDNDIANVVRKL	63
QY	74	YATGNFSDQVYVHQEGRIIYQVTERPLIAEINFEGNRLPKKEGLQELKNAGLAVGOPLK	133
Db	64	FLSQGVDDVKASREGNTLVTVVMPKPVISNVVIVGNKSIDPEAIKONLDANGFKVGVDVNL	123
QY	134	QATVQMIETELTNQYISQGYNTEITVVKOTMLDGNRVKLDMTFAECPARVVDINIIGNQ	193
Db	124	RAKLEBFRKGIVEHYNSVGRYNAKVDAIVNTLPNNSAETKIQINEDDVALFKREITFEGNE	183
QY	194	HFSADLIDVLATKDNKINPLSKADRYTOEKLTVSLENRAKYLNAFGRFEETKDAKLN	253
Db	184	AFSSGKLADQOMELQTSWNKLF-GNEFDQTOFNKDLTSLRSYVLDGRYQAQFQILDDTVKL	242
QY	254	NEDKNRIFVEISLHCEGPOVRFQTOFLGNL-TYQAELEALLKFAEAG-FSQAOMEQTT	311
Db	243	SDDKKEPCL-IS-EEGDLTV-KTRVSGMGWGSNAELAPILETIOLNGLFRFTSVLEVE	299
QY	312	NNISTKFGDDGYIYAQIRPVTRINDESRVDVEYYIDPVHVVRVINFTEGNFKTODEVL	371
Db	300	QRNKSXLGERGYATAQVNVHPTFDEQDKTISLDFIVEACKSYTVROIREGNTSSADSTL	359
QY	372	REMRQLEGALASNOKIQISRLARLMTGTFKHVTVTRVPNSPDQDVNVFVEEQPSGS	431
Db	360	ROEMROEGAWJUSSELVELKULRDTGYFESVETKTEAIPGS-DQDVYIYKVKERNTGS	418
QY	432	STIAAGYSOGGVTFQDFYSONNFMGTGKHVNASFRRSETREYISUGMTPNFTVTVNGVSO	491
Db	419	INFGIGYGTESGLSYOASIKQDNFLGMGSSISLGGTRNDYCTIVNLGYNEPFTTKDGV--	476
QY	492	SLSGYRRTKYD---NKNTISNVLDYSGSLSYGVPIDENQRIISFGLNADNTKLHGGREF	548

RESULT 13

```
Db 477 SLGGNVSEEDSSKSNSTASGYGRTSYGNNLTGLPPVNNNSYYLGVTYNNKL----- 530
Qy 549 GISNV-KOLMADGGKIQVNDNGIPDFK-HDYTTYNAIILGWYSSLDPRVFTQGHSHVD 606
Db 531 --KNTAPEYNDRLYQSMKYNDSWTFKSHD--FDLSFGWYNSLNRGYFPTKGVYRANIG 585
Qy 607 --LTVFGDKTHQVYVYOGNYIRPP-----IKSVLRGVAKILGYGNLPPFYENF 653
Db 586 GRVTIPGSDNYYKLNABAOGFYPLDREHGWVLSRSSASPADGFS---GKRLPFYQY 641
Qy 654 YAGGVYRGVYDQSSGLGRSQAYLTARGQQTTL--GEVVGSNALATFGSELILPLPKG 711
Db 642 SAGGIGSURGAYGAGIGNA-IYTRQCPDSYCLVSSVDVIGNAMVTASTELIVPTPVA 700
Qy 712 DW-IDQVRPIFIEGGVDFDTGMDKQIDLTQFKDPQATAEQNAKAA--NRPLLTQDKQ 768
Db 701 DKNQSVRTSLFVDAASVWNT-----RWKAEDRAKFAKLVNPDYDSFR 744
Qy 769 LRYSGVGTATWTPIGPLISYAKPLNKKQNDQDTVQFGSVF 813
Db 745 VRASGVALQWQSPITGLVFSYAKPLKRYQGDIEQFQFSGTGF 789

RESULT 14
Q9ZE03
ID Q9ZE03 PRELIMINARY; PRT; 768 AA.
AC Q9ZE03;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE OUTER MEMBRANE PROTEIN OMP1 (OMP1).
GN RPI60.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=792;
RN [1]
RC STRAIN=MADRID E;
RP SEQUENCE FROM N.A.
RX MEDLINE=90039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria.";
RL Nature 396:133-140(1998).
DR EMBL: AJ235270; CAAL4627.1; -
DR InterPro; IPR00184; Bac_surfaG_d15.
DR Pfam; PF01103; Bac_surfaceAg; 1.
KW Complete proteome.
SQ SEQUENCE 768 AA; 87196 MW; BIAB12D783D9FDCB CRC64;

Query Match 17.98; Score 754; DB 16; Length 768;
Best Local Similarity 26.38; Pred. No. 2.1e-33;
Matches 216; Conservative 145; Mismatches 383; Indels 76; Gaps 20;

Qy 7 KGFOVSAMTAVMVMSVTHAAQAFMANDITITGLQRTVIESLOSVPFLRGVYSENLQ 66
Db 2 KIISIKTLILLIFVYHIFADVIKIKITIEGNHRVERSTIESYLNKLVGEYNNKS 61
Qy 67 ADGVKALYATNFSDVQVY-HQEGRIIVQVTERPLIAINEGNRLIPKEGLQGLKNAG 125
Db 62 DEATKRLVATSLFRININMYITDGNLVNVTETPISSVWFSGSKIKTNILAKEIYMS 121
Qy 126 LAVGQPLQAQVOMIETELTNQYISQGYNTIETVQTMLDGNRVKLDMTFAEKGPARV 185
Db 122 ---GESLSQAQIELDVKKILEIYRSGRFSTKVPKIKSLNNRVKVFIDFAEGPKTVIK 178
Qy 186 DINIIGNOHFSDADLIDVLATKDNK-INPLSKADRYTOEKLVTSLNRAKYLNAGVRF 244
Db 179 SIYSGNGEHSDELKSVLTKESRWFRLESNDTIDPRVEYDKELLREFYQSVGFAD 238
```

```
Qy 245 EIKDAKLINEDKNRIFVEISLHEGEQVRFQGTQFLGNLTYYTQ-AELEALLKFKAEEGFS 303
Db 239 RVISASVALNDTKKEYTFTTYSIEGEKYRFGNVTIDNKLNTINIKQLNIVNIKQCKIEN 298
Qy 304 QAMLEQTTNNISTKGGDDGYAAQIRPVTRINDESTRVDVEYIIDPVHPVYVRRINFTCN 363
Db 299 MKTVDDIAEKIGEYFTANGYPANVYPIDIKND-NHTADIKFPIIEKADKVYINKNINNN 357
Qy 364 FKTDQEVLRREMRLEGALASNOIKQLSRARLMRTGFFKHVTVDTRPVNSPDQVDNV 423
Db 358 LKTEDHVIIRAFKIEEGDVWNRSYIEKGERNLRLNIDFEKVSISLAQT-KADKKYDVNVE 416
Qy 424 VEEQPSGSTIAAGYSQSGGVTFQDVSONPMGCKHKHVNASFSESTREYVSLGWTNPY 483
Db 417 VDEKSTSTGFDLGYNTAGGLFGRESFLERNLVGTGKLILNAGVQVSKNSTSYGCTIDPH 476
Qy 484 FTVNGVSQSLSGY---RKTKYDNKNISNYVLDYSVGSLSYGPIDE--NQRISEGLNA 537
Db 477 FLDRDLSUSVNAFRNYTGRGASVLTNTDQSYKLSHSGVKSIGYDMKEDLSHEIDYLIKR 536
Qy 538 DNTKLHGRFRMGISNVKQLMADGGKIQVNDNGIPDFKHDTYTYNAILGNWYSSLDPRV 597
Db 537 DILSAPSP-----SNSIFLNEQMGKLITSAIG-----HTIT-----YDQDNKIYP 577
Qy 598 TQG--MSHSVLTVGFGDKTHQVYVYQGYNIYRPFIKKSVLRGYAKLG-----YGNLPP 649
Db 578 KNGYLVSGTQEFAGVGGDNKYIKHEIECKFYKSFNNKVTCLKLSAAGDMAGLGKMWRI 637
Qy 650 YENFYAGGVSGVYDQSSGLGRSQAYLTARGQQTTLGEEVVGGNALATFGSELILPLPF 709
Db 638 SDRFNLGDY-SLRGFASGGVGRP-----EKNTNEGLGGERYYTFTSELNFPPTPV 685
Qy 710 KGDIDQVRPVIFIEGGVDFDTGMDKQIDLTQFKDPQATAEQNAKAAANRPLLTDQDKOL 769
Db 686 PEEF--NFTGAVFIDLGSVWG-VGLNKK-----QYKTPNG-----FYNDQSL 724
Qy 770 RYSAGVGATWTPIGPLISYAKPLNKKQNDQDTVQFQI 809
Db 725 RASVGFGFIWVTRFAPIRMDWGFPIKKQYD--DTQNFHL 762

RESULT 15
Q92Q48
ID Q92Q48 PRELIMINARY; PRT; 776 AA.
AC Q92Q48;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE PUTATIVE OUTER MEMBRANE TRANSMEMBRANE PROTEIN.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RC STRAIN=1021;
RP SEQUENCE FROM N.A.
RX MEDLINE=21368234; PubMed=11474104;
RA Galibert F., Finan T.M., Long S.R., Puehler A., Abola P., Ampe F.,
RA Barloy-Hubler F., Barnett M.J., Becker A., Boistard P., Bothe G.,
RA Boutry M., Bowser L., Buhrmester J., Cadieu E., Capela D., Chain P.,
RA Cowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F.,
RA Gloux S., Godrie T., Goffeau A., Golding B., Gouzy J., Gurjal M.,
RA Hernandez-Lucas I., Hong A., Huizar L., Hyman R.W., Jones T., Kahn D.,
RA Kahn M.L., Kalman S., Keating D.H., Kiss E., Komp C., Lelaure V.,
RA Masuy D., Palm C., Peck M.C., Pohl T.M., Portetelie D., Purnelle B.,
RA Ramsperger U., Surzycki R., Thebault P., Vandenbol M.,
RA Vothoelster F.J., Weidner S., Wells D.H., Wong K., Yeh K.-C., Batut J.;
RT "The composite genome of the legume symbiont Sinorhizobium meliloti.";
RL Science 293:668-672(2001).
DR EMBL: AL591787; CAC46081.1; -
KW Complete proteome.
SQ SEQUENCE 776 AA; 84520 MW; CE261E6372A94726 CRC64;
```



[illegible]

Search completed: September 5, 2002, 10:02:56  
Job time: 530 sec





Db 301 gfsqamleqttnnistkfgdgyyaaqrprvtrindesrtvdveyyidpvhpyvrrinf 360  
 QY 361 TGFKTKQDEVLREMRLEGALASNQIKLSRARLMRTGFFKHVTVDTRVPVNSPDQVDV 420  
 Db 361 tgnfktqdevlrremrqlqegalsnqkqlsrlarlmrtgffkhvtvdtrvpvnsdpdqv 420  
 QY 421 NFVVEQPSGSSTIAAGYSQSGGVTFQFQVSYQNNFMGTGKHVNASFSRSETREYISLGMT 480  
 Db 421 nfvveeqpsgstiaagysqsggvtfqfqsqnnfmgtgkhvnasfsrsetrevyslgmt 480  
 QY 481 NPFTVNGVSQSLSGYRKTKYDNKNISNVYVLSYSGSLSYGYPIDBNQRIISFGLNADNT 540  
 Db 481 npftvngvsqslsgyrrtktdnknisnvyvlsygsysgyypidenqrisfqlnadnt 540  
 QY 541 KLHGGRFMGTSNVKQJLMDADGKIQVDNNGIPDFKHDTTYNAILGWNYSLSLDRPVPTQG 600  
 Db 541 klhggrfmgtisnvkqlmdadgkqlvdnngipdfkhdtytnailgwnyslsldrpvptqg 600  
 QY 601 MSHSVDLTVGFGDKTHQVYVYQGNIRPFIFIKKSVLRGYAKLGXGNLPPFYENFYAGCYGS 660  
 Db 601 mshsvdltvfgdkthqkvvyqgnirpfifikksvlrgyaklgxgnlppfyenyfagcygs 660  
 QY 661 VRGYDSSLGPRQAYLTARRGQQTTLGEVVGGNALATFGSELILPLPFGDWIDQVRPV 720  
 Db 661 vrgydgsslgprsqayltarrgqqtllgevvvggnalatfgselilplpfkgdwldqvrpv 720  
 QY 721 IFTEGGQVFTTGMKDQITDLTQFKDQATAEQNAKAANRPLLTQDKOLRYSGAGVATWY 780  
 Db 721 ifteggqvfttgmkdqitdltfkdqataeqnakaanrplltqdkqlrysgagvatwy 780  
 QY 781 TPIGPLSISYAKPLNKKQNDQDTVQFQIGSVF 813  
 Db 781 tpigplsisyakplnkkqndqdtvqfqigsvf 813

RESULT 2

AAI44391  
 ID AAY44391 standard; Protein; 813 AA.  
 AC AAY44391;  
 XX  
 DT 14-MAR-2000 (first entry)  
 XX  
 DE M. catarrhalis (ATCC 43617) BASB027 polypeptide #2.  
 XX  
 KW BASB027: OMP85; outer membrane protein; otitis media; treatment;  
 KW diagnosis; bacterial infection.  
 XX  
 OS Moraxella catarrhalis.  
 PN WO9963093-A2.  
 XX  
 PD 09-DEC-1999.  
 XX  
 PF 31-MAY-1999; 99WO-EP03822.  
 XX  
 PR 03-JUN-1998; 98GB-0011945.  
 PR 08-MAR-1999; 99GB-0005304.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 XX  
 PI Vinalis-Bassols C;  
 XX  
 DR WPI: 2000-105700/09.  
 DR N-PSDB; AA29551.  
 XX  
 PT Novel BASB027 polynucleotide and polypeptides from Moraxella  
 PT catarrhalis useful for treating M. catarrhalis infection such as otitis  
 PT media  
 XX  
 PS Claim 1; Page 102-104; 109pp; English.  
 XX  
 CC The present sequence is BASB027 polypeptide, which shows significant

CC homology to Neisseria meningitidis OMP85 outer membrane protein. It is  
 CC encoded by DNA obtained from chromosomal DNA library of Moraxella  
 CC catarrhalis strain Mc2931 (ATCC 43617). BASB027 polynucleotide and  
 CC polypeptide can be used for diagnosis and staging of disease, determining  
 CC susceptibility to a disease and to prepare medicaments for treating M.  
 CC catarrhalis infections, especially otitis media. The BASB027 DNA can be  
 CC used as probe for screening of genetic mutations, serotype, taxonomic  
 CC classification or identification. BASB027 agonists, antagonists and  
 CC antibodies may be used to prevent and/or treat bacterial infections.  
 XX  
 SQ Sequence 813 AA;

Query Match 92.1%; Score 749; DB 21; Length 813;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 749; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 QLADGVKALYATGNFSDVQVYHQEGRIYQVTERPLIAEINFEGNRLIPKEGLQGLKNA 124  
 Db 65 qladgvkalyatgnfsvdvyvqhgegrliyqvrterpliaeinfegnrlipkeglqglkna 124  
 QY 125 GLAVGQPLKQATVQMIETELTNQYISQGYNTETITVQTMLDGNRVKLDMTFAEGKPARV 184  
 Db 125 glavgqplkqatvqmieteltnqyisqgyntetitvqtmldgnrvkldmtfaegkparv 184  
 QY 185 VDINIIGNOHFSDADLDVLAIDKKNIPLSKADRYTOEKLVTSLLENRAKYLNAGFVRF 244  
 Db 185 vdnignohfstdadldvlaidkkniplskadytqeklvtslenlrakylnagfvrf 244  
 QY 245 EIKDAKUNINEDKNRIFVEISLHGEQYRFGQTQFLGNLYTQAELEALUKFAKEEGFSQ 304  
 Db 245 eikdakuninedknrifveislhegeqyrfgqtqflgnlytqaeleallkfkaeegfsq 304  
 QY 305 AMLEQTTNNISTKFGDDGYVYAOIRPVTRINDESRTVDVEYIDVHPVYVRRNFTGNF 364  
 Db 305 amleqttnnistkfgddgyvyaqirpvtrindesrtvdveyyidvpvpyvrrnftgnf 364  
 QY 365 KTQDEVLRREMRQLEGALASNQIKLSRARLMRTGFFKHVTVDTRVPVNSPDQVDVAVFW 424  
 Db 365 ktqdevlrremrqlqegalsnqkqlsrlarlmrtgffkhvtvdtrvpvnsdpdqvavfw 424  
 QY 425 EEPQSGSTIAAGYSQSGGVTFQFQVSYQNNFMGTGKHVNASFSRSETREYISLGMTNYPF 484  
 Db 425 eepqsgstiaagysqsggvtfqfqsqnnfmgtgkhvnasfsrsetrevyslgmtnypf 484  
 QY 485 TVNGVSQSLSGYRKTKYDNKNISNVYVLSYSGSLSYGYPIDBNQRIISFGLNADNTKLHG 544  
 Db 485 tvngvsqslsgyrrtktdnknisnvyvlsygsysgyypidenqrisfqlnadntklhg 544  
 QY 545 GRFMGISNVKQLMADGKIQVDNNGIPDFKHDTTYNAILGWNYSLSLDRPVFPPTQGMSHS 604  
 Db 545 grfmgisnvkqlmadgkqvvdnngipdfkhdtytnailgwnyslsldrpfvptqgmshs 604  
 QY 605 VDLTVGFGDKTHQVYVYQGNIRPFIFIKKSVLRGYAKLGXGNLPPFYENFYAGGYSGVRGY 664  
 Db 605 vdltvfgdkthqkvvyqgnirpfifikksvlrgyaklgxgnlppfyenyfaggysgvrgy 664  
 QY 665 DQSSSLGPRQAYLTARRGQQTTLGEVVGGNALATFGSELILPLPFGDWIDQVRPVFIE 724  
 Db 665 dqssslgprsqayltarrgqqtllgevvvggnalatfgselilplpfkgdwldqvrpvfie 724  
 QY 725 GGQVFDFTTGMKDQITDLTQFKDQATAEQNAKAANRPLLTQDKOLRYSGAGVATWYPIG 784  
 Db 725 ggqvfdfttgmkdqitdltfkdqataeqnakaanrplltqdkqlrysgagvatwypig 784  
 QY 785 PLSISYAKPLNKKQNDQDTVQFQIGSVF 813  
 Db 785 plsisyakplnkkqndqdtvqfqigsvf 813

Search completed: September 5, 2002, 10:13:58  
 Job time: 386 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 10:14:27 ; Search time 25.27 Seconds  
(without alignments)  
785.833 Million cell updates/sec

Title: US-09-701-711-4  
Perfect score: 813  
Sequence: 1 MRNSYKGFQVSAMTAVMM.....LNKKQNDQDTVQFQIGSVF 813

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 231628 seqs, 2442594 residues

Word size : 20

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued\_Patents\_AA:\*

- 1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*
- 2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*
- 3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*
- 4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*
- 5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*
- 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length	ID Description
-----			

No matches found

Search completed: September 5, 2002, 10:14:27  
Job time: 285 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 10:15:14 ; Search time 39.83 Seconds  
(without alignments)  
1961.351 Million cell updates/sec

Title: US-09-701-711-4  
Perfect score: 813  
Sequence: 1 MRNSYFKGFQVSAMTAVMM.....LNKKQNDQDTDTVQFQIGSVF 813

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size : 20

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : PIR\_71.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
-----						

No matches found

Search completed: September 5, 2002, 10:15:14  
Job time: 242 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 10:22:07 ; Search time 20.99 Seconds  
(without alignments)  
1499.714 Million cell updates/sec

Title: US-09-701-711-4  
Perfect score: 813  
Sequence: 1 MRNSYFKGFQVSAMTMVM.....LNKKQNDQDQTVQFQIGSVF 813

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 105224 segs, 38719550 residues

Word size : 20

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	%
No.	Score Match Length DB ID	Description
-----		

No matches found

Search completed: September 5, 2002, 10:22:07  
Job time: 454 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 10:21:40 ; Search time 63.52 Seconds  
(without alignments)  
2214.183 Million cell updates/sec

Title: US-09-701-711-4  
Perfect score: 813  
Sequence: 1 MRNSYKFGQVSAMTAVMM.....LNKKQNDQDTVQFQIGSVF 813

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size : 20

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

- Database : SPTREMBL\_19:\*
- 1: sp\_archaea:\*
  - 2: sp\_bacteria:\*
  - 3: sp\_fungi:\*
  - 4: sp\_human:\*
  - 5: sp\_invertebrate:\*
  - 6: sp\_mammal:\*
  - 7: sp\_mhc:\*
  - 8: sp\_organelle:\*
  - 9: sp\_phage:\*
  - 10: sp\_plant:\*
  - 11: sp\_rodent:\*
  - 12: sp\_virus:\*
  - 13: sp\_vertebrate:\*
  - 14: sp\_unclassified:\*
  - 15: sp\_rvirus:\*
  - 16: sp\_bacteriap:\*
  - 17: sp\_archheap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
-----					

No matches found

Search completed: September 5, 2002, 10:21:40  
Job time: 457 sec



OM protein - protein search, using sw model

Run on: September 5, 2002, 10:04:37 ; Search time 61.46 seconds  
(without alignments)  
1469.297 Million cell updates/sec

Title: US-09-701-711-4  
Perfect score: 813  
Sequence: 1 MRNSYFKGFQVSAMTMMVM.....LNKKQNDQDVTQVQIGSVF 813

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Word size : 15

Total number of hits satisfying chosen parameters: 4

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :			
A_Geneseq_032802.*			
1:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*		
2:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*		
3:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*		
4:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*		
5:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*		
6:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*		
7:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*		
8:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*		
9:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*		
10:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*		
11:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*		
12:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*		
13:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*		
14:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*		
15:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*		
16:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*		
17:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*		
18:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*		
19:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*		
20:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*		
21:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*		
22:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match	Description
1	813	100.0	813 21 AAY44390 M. catarrhalis BAS
2	749	92.1	813 21 AAY44391 M. catarrhalis (AT
3	16	2.0	17 21 AAY44393 Moraxella catarrha
4	15	1.8	16 21 AAY44392 Moraxella catarrha

ALIGNMENTS			

ID	AA44390 standard; Protein; 813 AA.
XX	
AC	AA44390;
XX	
DT	14-MAR-2000 (first entry)
XX	
DE	M. catarrhalis BASB027 polypeptide #1.
XX	
DE	BASB027; OMP85; outer membrane protein; otitis media; treatment;
KW	diagnosis; bacterial infection.
XX	
OS	Moraxella catarrhalis.
XX	
PN	WO9963093-A2.
XX	
PD	09-DEC-1999.
XX	
PF	31-MAY-1999; 99WO-EP03822.
XX	
PR	03-JUN-1998; 98GB-0011945.
XX	
PR	08-MAR-1999; 99GB-0005304.
XX	
XX	(SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
PA	
XX	
PI	Vinals-Bassols C;
XX	
XX	
DR	WPI: 2000-105700/09.
DR	N-PSDB; AA29550.
XX	
PT	Novel BASB027 polynucleotide and polypeptides from Moraxella
PT	catarrhalis useful for treating M. catarrhalis infection such as otitis
PT	media -
XX	
PS	Claim 1; Fig 3; 109pp; English.
XX	
CC	The present sequence is BASB027 polypeptide, which shows significant
CC	homology to Neisseria meningitidis OMP85 outer membrane protein. It is
CC	encoded by DNA obtained from chromosomal DNA library of Moraxella
CC	catarrhalis strain Mc2931 (ATCC 43617). BASB027 polynucleotide and
CC	polypeptide can be used for diagnosis and staging of disease, determining
CC	susceptibility to a disease and to prepare medicaments for treating M.
CC	catarrhalis infections, especially otitis media. The BASB027 DNA can be
CC	used as probe for screening of genetic mutations, serotype, taxonomic
CC	classification or identification. BASB027 agonists, antagonists and
CC	antibodies may be used to prevent and/or treat bacterial infections.
XX	
SQ	Sequence 813 AA;
Query Match 100.0%; Score 813; DB 21; Length 813;	
Best Local Similarity 100.0%; Pred. No. 0;	
Matches 813; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 MRNSYFKGFQVSAMTMMVMVMSHQAQADFMANDITGLQRTVTSQSVPFLRGQV 60
Db	
Db	1 mnsyfkqfqsamtmavmmvmshtaqaadfmnditglqrvtiesqsvlpflrgqv 60
QY	61 VSENLADGKVKALYATGNFSDQVYHQEGRYIVQVTERPLIAEINFEGRNLPKEGLQEG 120
Db	
Db	61 vsenladgkvkalyatgnfsdqvyhqegrilyqvterpliaeinfeqrnlipkeglqeg 120
QY	121 LKNAGLAVGQPLKQATVQMIETELTNQYISQGYNTEITVTKOTMLDGNRVKLDMTFAEGK 180
Db	
Db	121 lknaglavgqplkqatvqmieteltnqyisqgynteitvktqldgnrvkldmtfaegk 180
QY	181 PARVVDINIIQNHQSFADLIDVLAIKDNKINPLSKADRYTQEKLVTSLENLRAKYLNAG 240
Db	
Db	181 parvvdiniinqhfsadlidvlaikdnkinplskadrytqeklvtslenlrakylnag 240
QY	241 FVRFELKDAKLINEDKNRIFVEISLHGEQYRFGOTQFLGNLTYYTQAELEALLKFAEE 300
Db	
Db	241 fvrfelkdaklinedknrifveislhgeqyrfgotqflgnlttytqaeleallkfaee 300

301 GFSQAMLEQTTNNISTKFGDDGYYIAQIRPVTRINDESRVTVDVEYYIDPVHPVYRRINF 360  
 301 gfsqamleqttnnistkfgddgyyaqirpvrtrindestrvtdveyidpvhpyrrinf 360  
 361 TGNPKTQDEVLRMRQLEGALASNOKIQLSRARLMRTGFFKHHVTVTRPVPNSPDQVDV 420  
 361 tgnpktqdevlrmerqlegalasnkqlsrrarlmrtgffkhhvtvtrpvpnsdpdqv 420  
 421 NFVVEEQSGSSTIAAGYSQSGGVTFQDVSONNFMGTGKHVNASFSESRSEYVSLGMT 480  
 421 nfveeqsgsstiaagysqsggvtfqdvsonnfmgtgkhvnasfse srse yvslgmt 480  
 481 NPYFTVNGVSQSLGYRKTXYDNKNISYVLDYVGGSLSYCYPIDENQRIISFGLNADNT 540  
 481 npyftvngvsqslgyrktxydnknisyvldyvggslsygypidenqrisf glnadnt 540  
 541 KLHGGRFMGINSVKOLMADGGKIQVDNNGIPDFKHDTYTYNAILGWNYSLLDRPVFTQ 600  
 541 klhggrfmgisnkvqlmadggkigvdnngipdfkhdttytnailgwnyslldr pvtq 600  
 601 MSHSVDLTVGFGDKTHOKVYVQGNTRYRPFIRKSVLRGVAKLGYGNLFPYENFVAGGYGS 660  
 601 mshsvdltvvgfkdthokvyvqgntryrpfirksvlrgvaki gygnlfp yenf vaggys 660  
 661 VRGYDQSSGLSPRSQAYLTARRGQOTTLGEVYVGGNALTGFGSELILPLPFKGDWIDQVRPV 720  
 661 vrgydqssglprsqayltarrgqottlgevvyggnaltfgselilplpfkgdwidqvrpv 720  
 721 IFIEGGQVFDTTGMDKQIFDLTQFKDPOATAEONAKAANRPLLTQDKOLRYSGAGVATWY 780  
 721 ifieggqvfdttgmdkqifdl tqfkdpqataeqnakaanrplltqdkolrysagvatwy 780  
 781 TPIGPLSISYAKPLNKKONDQTDVTFQIGSVF 813  
 781 tpigplsisyakplnkkndqtdvtvfqigsvf 813

RESULT 2

AA44391  
 ID AA44391 standard; Protein; 813 AA.  
 XX  
 AC AA44391;  
 DT 14-MAR-2000 (first entry)  
 DE  
 KW M. catarrhalis (ATCC 43617) BASB027 polypeptide #2.  
 KW BASB027; OMP85; outer membrane protein; otitis media; treatment;  
 KW diagnosis; bacterial infection.  
 OS Moraxella catarrhalis.  
 PN W09963093-A2.  
 XX  
 PD 09-DEC-1999.  
 XX  
 PF 31-MAY-1999; 99WO-EP03822.  
 XX  
 PR 03-JUN-1998; 98GB-0011945.  
 PR 08-MAR-1999; 99GB-0005304.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 XX  
 PI Vinals-Bassols C;  
 XX  
 DR WPI; 2000-105700/09.  
 DR N-PSDB; AAZ29551.  
 XX  
 PT Novel BASB027 polynucleotide and polypeptides from Moraxella  
 PT catarrhalis useful for treating M. catarrhalis infection such as otitis  
 PT media  
 XX  
 PS Claim 1; Page 102-104; 109pp; English.

XX The present sequence is BASB027 polypeptide, which shows significant  
 CC homology to Neisseria meningitidis OMP85 outer membrane protein. It is  
 CC encoded by DNA obtained from chromosomal DNA library of Moraxella  
 CC catarrhalis strain Mc2931 (ATCC 43617). BASB027 polynucleotide and  
 CC polypeptide can be used for diagnosis and staging of disease, determining  
 CC susceptibility to a disease and to prepare medicaments for treating M.  
 CC catarrhalis infections, especially otitis media. The BASB027 DNA can be  
 CC used as probe for screening of genetic mutations, serotype, taxonomic  
 CC classification or identification. BASB027 agonists, antagonists and  
 CC antibodies may be used to prevent and/or treat bacterial infections.  
 XX  
 SQ Sequence 813 AA;

Query Match 92.1%; Score 749; DB 21; Length 813;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 749; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 65 QLADGVKALYATGNFSDVYVHQEGRIYQVTERPLIAEINFEGRNLRIPKEGLQELKNA 124  
 Db 65 qladgvkalyatgnfsdvvyvqhgrliyvterpliaeinfegnrlipkeglqelkna 124  
 Qy 125 GLAVGQPLKQATVQMIEETELNQYISQGYNTEITVKTQMLDGNVTKLDMTEAGKPARV 184  
 Db 125 glavgqpikqatvqmietelntqyisqgynteitvktquml dgnrvtkl dmtfaegkparv 184  
 Qy 185 VDINIIGNHSDADLIDVLAIKDNKINPLSKADRYTQEKLVTSLENRAKYLNAGFVR 244  
 Db 185 vdiniignhgsdadlidvlaikdnkinplskadrytqeklvtslenrakylnagfvrf 244  
 Qy 245 EIKDAKLINEDKNRIFVEISLHEGEQYRFGQTQFLGNLTYYTQAELEALLKFAEGFSQ 304  
 Db 245 eikdaklinedknrifveislhegeqyrfgqtqflgnlt ytgaeleallkfkaegfsg 304  
 Qy 305 AMLEOTTNNISTKFGDDGYYIAQIRPVTRINDESRVTVDVEYYIDPVHPVYRRINFTGNF 364  
 Db 305 amleottnnistkfgddgyyaqirpvrtrindestrvtdveyidpvhpyrrinf tgnf 364  
 Qy 365 KTQDEVLRMRQLEGALASNOKIQLSRARLMRTGFFKHHVTVTRPVPNSPDQVDV 424  
 Db 365 ktqdevlrmerqlegalasnkqlsrrarlmrtgffkhhvtvtrpvpnsdpdqv dnv 424  
 Qy 425 EQPSGSSSTIAAGYSQSGGVTFQDVSONNFMGTGKHVNASFSESRSEYVSLGMTNYPF 484  
 Db 425 eqpsgsstiaagysqsggvtfqdvsonnfmgtgkhvnasfse srse yvslgmtncpyf 484  
 Qy 485 TVNGYSQSLGYRKTXYDNKNISYVLDYVGGSLSYCYPIDENQRIISFGLNADNTKLHG 544  
 Db 485 tvngysqslgyrktxydnknisyvldyvggslsygypidenqrisf glnadntklhg 544  
 Qy 545 GRFMGISNVKQLMADGGKIQVDNNGIPDPKHDTYTYNAILGWNYSLLDRPVFTQGMSHS 604  
 Db 545 grfmgisnkvqlmadggkigvdnngipdpkhdttytnailgwnyslldr pvtqgmshs 604  
 Qy 605 VDLTVGFGDKTHOKVYVQGNTRYRPFIRKSVLRGVAKLGYGNLFPYENFVAGGYSGVRG 664  
 Db 605 vdltvvgfkdthokvyvqgntryrpfirksvlrgvaki gygnlfp yenf vaggysvrg 664  
 Qy 665 DQSSILGPRSOAYLTARRGQOTTLGEVYVGGNALTGFGSELILPLPFKGDWIDQVRPV 724  
 Db 665 dqssilgprsqayltarrgqottlgevvyggnaltfgselilplpfkgdwidqvrpv 724  
 Qy 725 GGOVFDTTGMDKQIFDLTQFKDPOATAEONAKAANRPLLTQDKOLRYSGAGVATWTPIG 784  
 Db 725 ggovfdttgmdkqifdl tqfkdpqataeqnakaanrplltqdkolrysagvatwtpig 784  
 Qy 785 PLISISYAKPLNKKONDQTDVTFQIGSVF 813  
 Db 785 plisisyakplnkkndqtdvtvfqigsvf 813

AA44393  
ID AAY44393 standard; peptide; 17 AA.

XX AC AAY44393;  
XX DT 14-MAR-2000 (first entry)  
XX DE Moraxella catarrhalis BASB027-specific synthetic peptide #2.  
XX KW BASB027; OMP85; outer membrane protein; otitis media;  
XX KW bacterial infection.  
XX OS Synthetic.

XX PN WO9963093-A2.

XX PD 09-DEC-1999.

XX PF 31-MAY-1999; 99WO-EP03822.

XX PR 03-JUN-1998; 98GB-0011945.  
XX PR 08-MAR-1999; 99GB-0005304.

XX PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
XX PI Vinals-Bassols C;

XX DR WPI; 2000-105700/09.

XX PT Novel BASB027 polynucleotide and polypeptides from Moraxella

XX PT catarrhalis useful for treating M. catarrhalis infection such as otitis  
XX PT media -  
XX PS Example 8; Page 64; 109pp; English.

XX CC The present sequence is a Moraxella catarrhalis BASB027

XX CC polypeptide-specific peptide which can be coupled to KLH and used for  
XX CC producing antibodies against BASB027. Anti-BASB027 antibodies can be used  
XX CC to treat bacterial infections.

XX SQ Sequence 17 AA;

Query Match 2.0%; Score 16; DB 21; Length 17;  
Best Local Similarity 100.0%; Pred. No. 1e-08;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 676 YLTARRGQQTTLGEVV 591

Db 1 yltarrgqgttlgevv 16

RESULT 4

AA44392

ID AAY44392 standard; peptide; 16 AA.

XX AC AAY44392;

XX DT 14-MAR-2000 (first entry)

XX DE Moraxella catarrhalis BASB027-specific synthetic peptide #1.

XX KW BASB027; OMP85; outer membrane protein; otitis media;  
XX KW bacterial infection.

XX OS Synthetic.

XX PN WO9963093-A2.

XX PD 09-DEC-1999.

XX PF 31-MAY-1999; 99WO-EP03822.

XX

PR 03-JUN-1998; 98GB-0011945.  
PR 08-MAR-1999; 99GB-0005304.

XX PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.

XX PI Vinals-Bassols C;

XX DR WPI; 2000-105700/09.

XX PT Novel BASB027 polynucleotide and polypeptides from Moraxella  
XX PT catarrhalis useful for treating M. catarrhalis infection such as otitis  
XX PT media -

XX PS Example 8; Page 64; 109pp; English.

XX CC The present sequence is a Moraxella catarrhalis BASB027

XX CC polypeptide-specific peptide which can be coupled to KLH and used for  
XX CC producing antibodies against BASB027. Anti-BASB027 antibodies can be used  
XX CC to treat bacterial infections.

XX SQ Sequence 16 AA;

Query Match 1.8%; Score 15; DB 21; Length 16;  
Best Local Similarity 100.0%; Pred. No. 1.1e-07;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 790 YAKPLNKKQNDQDT 804

Db 2 yakplnkkqndqdt 16

Search completed: September 5, 2002, 10:04:38  
Job time: 511 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 10:05:08 ; Search time 25.36 Seconds  
(without alignments)  
783.044 Million cell updates/sec

Title: US-09-701-711-4  
Perfect score: 813  
Sequence: 1 MRNSYFKGFQVSAMTMAVM.....LNKKONDQDTVQFQIGSVF 813

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 231628 seqs, 24425594 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued\_patents\_AA:\*

- 1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*
- 2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*
- 3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*
- 4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*
- 5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*
- 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score Match	Length	ID	Description
---------------	----------------------	--------	----	-------------

No matches found

Search completed: September 5, 2002, 10:05:08  
Job time: 316 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 10:05:54 ; Search time 39.77 Seconds  
(without alignments)  
1964.310 Million cell updates/sec

Title: US-09-701-711-4  
Perfect score: 813  
Sequence: 1 MRNSYFKGFQVSAMTMAVMM.....LNKKQNDQTDTVQFQIGSVF 813

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : PIR\_71:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
---------------	-------	----------------	--------	----	----	-------------

No matches found

Search completed: September 5, 2002, 10:05:54  
Job time: 167 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 10:12:48 ; Search time 20.98 Seconds  
(without alignments)  
1500.429 Million cell updates/sec

Title: US-09-701-711-4  
Perfect score: 813  
Sequence: 1 MRNSYFKGFQVSAMTMAVM.....LNKKQNDQDTVQFQIGSVF 813

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score Match	Length	DB ID	Description
-----				

No matches found

Search completed: September 5, 2002, 10:12:48  
Job time: 411 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 10:12:20 ; Search time 63.51 Seconds  
(without alignments)  
2214.531 Million cell updates/sec

Title: US-09-701-711-4  
Perfect score: 813  
Sequence: 1 MRNSYFKGFQVSAMTMAVMM.....LNKKQNDQDTVQFQIGSVF 813

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

- Database : SPTREMBL19:\*
- 1: sp\_archaea:\*
  - 2: sp\_bacteria:\*
  - 3: sp\_fungi:\*
  - 4: sp\_human:\*
  - 5: sp\_invertebrate:\*
  - 6: sp\_mammal:\*
  - 7: sp\_mhc:\*
  - 8: sp\_organelle:\*
  - 9: sp\_phage:\*
  - 10: sp\_plant:\*
  - 11: sp\_rodent:\*
  - 12: sp\_virus:\*
  - 13: sp\_vertebrate:\*
  - 14: sp\_unclassified:\*
  - 15: sp\_rvirus:\*
  - 16: sp\_bacteriap:\*
  - 17: sp\_archheap:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES			
Result	Query		
No.	Score	Match Length DB ID	Description
-----			

No matches found

Search completed: September 5, 2002, 10:12:20  
Job time: 428 sec





7.04  
2.02  
2.4

39-19

June







## seq\_documentation\_block:

LOCUS AK003990 1629 bp mRNA linear HTC 19-JAN-2002  
 DEFINITION Mus musculus 18 days embryo whole body cDNA, RIKEN full-length enriched library, clone:1110030L07:homolog to PROTEIN CGI-51, full insert sequence.

## ACCESSION

AK003990.1 GI:12834981

## KEYWORDS

HTC; CAP trapper.

## SOURCE

Mus musculus (strain:C57BL/6J) 18 days embryo cDNA to mRNA,

clone\_lib:RIKEN full-length enriched mouse cDNA library

clone:1110030L07.

## ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

1 (sites)

Carninci,P. and Hayashizaki,Y.

High-efficiency full-length cDNA cloning

Meth. Enzymol. 303, 19-44 (1999)

99279253

PUBMED

10349636

## REFERENCE

2 (sites)

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,

Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

2049374

PUBMED

11042159

## REFERENCE

3 (sites)

Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,

Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,

Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,

Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,

Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Matsuoka,M.,

Toneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,

Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer

Genome Res. 10 (11), 1757-1771 (2000)

20530913

PUBMED

11076861

## REFERENCE

4 (sites)

The RIKEN Genome Exploration Research Group Phase II Team and the

FANTOM Consortium.

Functional annotation of a full-length mouse cDNA collection

Nature 409, 685-690 (2001)

5 (bases 1 to 1629)

Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,

Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,

Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,

Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,

Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,

Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,

Kurthara,C., Matsuyama,T., Miyazaki,A., Nishikawa,K., Nomura,K.,

Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,

Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,

Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,

Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,

Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I.,

Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and

Hayashizaki,Y.

Direct Submission

Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of

Physical and Chemical Research (RIKEN), Laboratory for Genome

Exploration Research Group, RIKEN Genomic Sciences Center (GSC),

RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,

Kanagawa 230-0045, Japan (E-mail:genome-res@gsr.riken.go.jp,

URL:http://genome.gsc.riken.go.jp/. Tel:81-45-503-9222,

Fax:81-45-503-9216)

Please visit our web site (http://genome.gsc.riken.go.jp/) for

further details.

cDNA library was prepared and sequenced in Mouse Genome

Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues. First strand cDNA was primed with a primer  
 [5' GAGAGAGAGCGCGCACTGAGTCTTTTCTTTTNN 3'], cDNA was  
 prepared by using trehalose thermo-activated reverse transcriptase  
 and subsequently enriched for full-length by cap-trapper. cDNA went  
 through one round of normalization to Rot = 5.0. Second strand cDNA  
 was prepared with the primer adapter of sequence[5'  
 GAGAGAGAGAGATCCAGAGCTCAATTAATTAATTAACCCGCCCC 3']. cDNA was  
 cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:  
 XhoI. Host: SOLR.

## FEATURES

source Location/Qualifiers  
 1..1629  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="MGI:1900625"  
 /db\_xref="taxon:10090"  
 /clone="1110030L07"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="18 days embryo"  
 misc\_feature 1..1629  
 /note="data source:SPTR, source key:09Y512, evidence:ISS  
 homolog to PROTEIN CGI-51"  
 BASE COUNT 376 a 392 c 484 g 377 t  
 ORIGIN

## alignment\_scores:

Quality: 204.00 Length: 521  
 Ratio: 0.761 Gaps: 24  
 Percent Similarity: 51.440 Percent Identity: 22.265

## alignment\_block:

US-09-701-711-2 x AK003990 ..

Align seg 1/1 to: AK003990 from: 1 to: 1629

334 ileAsnAspGluSerArgThrValAspValGlu.....Tyrty 346

151 CTGGGGGAGGAGCGGAGTTTGTGGNAGTTGAGCGGAGCGGAGGACGA 200

346 rleAspProValHisProValTyrValArgArgIleAsnPheThrGly 363

201 AATCTCGGAAAAAAGATGTGTGTTTCAGCATGTTTTCATTTGATGGAC 250

363 snPheLysThrGlnAspGluValLeuArgArgGluMetArgGlnLeu... 378

251 TTGGCGGAGCTAAGGATGACATCATCTTTGTGAAATCGGAGAGGTCTTT 300

379 .....GluGlyAlaLeuAlaSerAsnGlnLysIleGlnLeuSerArgAl 393

301 AAGGCTAAAAAACCCTCATTCAGGTAAATCGCGCATCTCATCAGCCGGGA 350

393 aArgLeuMetArgThrGlyPhePheLysHisValThrVal.....AspT 408

351 AAAACTGTTCTCGCTAGGAATTTTAGACAGTGGATGTTTTCATCGATA 400

408 hrArgProValProAsnSer....ProAspGlnValAspValAsnPheVal 423

401 CATGTCATGTTGAGTAAAGTCCCTGCGGCAATGGGTAGATGTCACCTTTGAA 450

424 ValGluGlu.....GlnProSerGlySerSerThrIleAlaAlaGlyTy 438

451 CTGACAGAGCTGAGGAGACTGACGGGCGAGTTACACACCATGTTGGAAA 500

438 rSerGlnSerGlyGlyValThrPheGlnPheAspValSerGlnAsnAsp 455

501 CAACGAAGGCGAGTATGGTA.....CTCGGCTCAAACTCCCAACC 541

455 heMetGlyThrGlyLysHisValAsnAlaSerPheSerArgSerGluThr 471

542 TTCTGGGACGACGAGAAAAAGTCACTTTCCAGTTTTTCTTATGGAACCAA 591

```
472 ArgGluValTyrSerLeuGlyMetThrAsnProTyr..... 483
   :::::::::::::::::::::
592 GAAATTCCTACGGCTGCTCTTCTCAAGCCACACGCTGGAACCTCGA 641
   :::::::::::::::::::::
484 .....PheThrValAsnGlyValSerGlnSerLeuSerGlyTyrTyr. 497
   :::::::::::::::::::::
642 GAGAAATTTCTCGTAAAC.....TTATATAAAGTTACTGGCGCAGTTCC 685
   :::::::::::::::::::::
498 .....ArgLysThrLysTyrAspAsnLysAsnIleSerAsn 509
   :::::::::::::::::::::
686 CGTGGAGCTCACTTCGGGAGACA..... 708
   :::::::::::::::::::::
510 TyrValLeuAspSerTyrGlyGlySerLeuSerTyrGlyTyrProIleAs 526
   :::::::::::::::::::::
709 .....GACAGAGAGTGCTGCAGAGTACAGTTTCCCTCTGTG 746
   :::::::::::::::::::::
526 pGluAsnGlnArg.IleSerPheGlyLeuAsnAlaAspAsnThrLysLeu 542
   :::::::::::::::::::::
747 GAAGACCAGTCACACTGTCAAGTGG.....AGGGTG 778
   :::::::::::::::::::::
543 HisGlyGlyArgPheMetGlyIleSerAsnValLysGlnLeuMetAlaAs 559
   :::::::::::::::::::::
576 sPTyrThrThrTyrAsnAlaIleLeuGlyTrpAsnTyr.SerSerLeuAs 592
   :::::::::::::::::::::
840 CACTCACTGAAGTCGTCTCTCGCATGCCATGGTCACTGACTCTCGAAA 889
   :::::::::::::::::::::
592 pArgProValPheProThrGlnGlyMetSerHisSerValAspLeuThrV 609
   :::::::::::::::::::::
890 TTCATCTATCTTGCCAAAGAGAGGGCCCTGTGTCAAAGTCAACGAGGAC 939
   :::::::::::::::::::::
609 al...GlyPhe.....GlyAspLysThrHisGlnLysValValTyrGln 622
   :::::::::::::::::::::
940 TGGCAGGCTACACTGGAGGAGATGTGAGCTTCATCAAGGAAGACTTTGAG 989
   :::::::::::::::::::::
623 GlyAsnIleTyrArgProPheIleLysLysSerValLeuArgGlyTyrAl 639
   :::::::::::::::::::::
990 CTTCACTGAATAAGCGCTGCGCTTGGACTCGGTATTCTCC.....AC 1033
   :::::::::::::::::::::
639 aLysLeuGlyTyrGlyAsnAsnLeuPro.....P 649
   :::::::::::::::::::::
1034 GTCTCTCTGGGTGGATCTGCTGTGCCATCGGTGACAGCCATCCAGCA 1083
   :::::::::::::::::::::
649 heTyrGluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAsp 665
   :::::::::::::::::::::
1084 TTGCTGACAGGTTTACCTGGGAGGCCCCACGAGTGTCCGAGGATTTAGC 1133
   :::::::::::::::::::::
666 GlnSerSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgl 682
   :::::::::::::::::::::
1134 ATGCAGCATTTGGACCCAGAGTGA..... 1160
   :::::::::::::::::::::
682 yGlnGlnThrThrLeuGlyGluValValGlyGlyAsnAlaLeuAlaThrP 699
   :::::::::::::::::::::
1161 .....GGAGATTACCTGGCGCGGCGGAGGCTACTGGGCTG 1194
   :::::::::::::::::::::
699 heGlySerGluLeuIleLeuProLeuProPheLys.....GlyAsp 712
   :::::::::::::::::::::
1195 GGGCGCTGCACCTCTACACCCACTGCCCTTCGGCCAGCGCAGGTTGCG 1244
   :::::::::::::::::::::
713 TrpIleAspGlnValArgProValIlePheIleGluGlyGlyGlnValPh 729
   :::::::::::::::::::::
1245 TTCGGAGAGCTTTTCAGAACTCACTTTTCTCCTCAATCGCGGCAACCTG.. 1292
   :::::::::::::::::::::
729 eAspThrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPheLysA 746
   :::::::::::::::::::::
1293 .....TGCACCTCACTATGTTGGAGG 1314
```

```
746 sProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLeu 762
   :::::::::::::::::::::
1315 GCCCAAAGCCCATATCCGGAAGCTAGCTGAGTGC..... 1349
   :::::::::::::::::::::
763 LeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrTr 779
   :::::::::::::::::::::
1350 .....ATCCGCTGCTCTATGAGCAGCGCTCGCTCT 1381
   :::::::::::::::::::::
779 p...TyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeuA 795
   :::::::::::::::::::::
1382 CGACTTGGCAACATCGCTCGCTGAGCTGAACCTACTGCATTCCTATGG 1431
   :::::::::::::::::::::
795 snLysLysGlnAsnAspGln...ThrAspThrValGlnPheGlnIleGly 810
   :::::::::::::::::::::
1432 GCGTGAGGGGGCGACAGAGTTTGATGGTGTCCAGTTTGGAGCTGGG 1481
   :::::::::::::::::::::
811 SerValPhe 813
   :::::::::::::::::::::
1482 ATTCGGTTC 1490
   :::::::::::::::::::::
seq_name: gb_gss:A0990340
seq_documentation_block: 636 bp DNA linear GSS 14-AUG-2000
LOCUS A0990340
DEFINITION Rfc011100 Photorhabdus luminescens strain W14 M13 library
           Photorhabdus luminescens genomic clone PLG01100, DNA sequence.
ACCESSION A0990340
VERSION A0990340.1 GI:9648934
KEYWORDS GSS.
SOURCE Photorhabdus luminescens.
ORGANISM Photorhabdus luminescens.
          Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
          Photorhabdus
REFERENCE 1 (bases 1 to 636)
AUTHORS ffrench-Constant,R.H., Waterfield,N., Burland,V., Perna,N.T.,
          Daborn,P.J., Bowen,D. and Blattner,F.R.
TITLE A genomic sample sequence of the entomopathogenic bacterium
JOURNAL Photorhabdus luminescens W14: potential implications for virulence
MEDLINE Appl. Environ. Microbiol. 66 (8), 3310-3329 (2000)
COMMENT 20378633
Contact: ffrench-Constant RH
Department of Biology and Biochemistry
University of Bath
South Building, Bath BA2 7AY, UK
Tel: (44) 1225 826621
Fax: (44) 1225 826779
Email: bsrfe@bath.ac.uk
This is one of 2,122 random reads from the M13 library. For
annotation of identified clones (BLASTX, BLASTN and mapping to E.
coli K12 genome) please see ffrench-Constant et al. 2000, Nucleic
Acids Res.
Seq primer: M13 Forward
Class: Shotgun.
FEATURES             Location/Qualifiers
     source            1..636
                     /organism="Photorhabdus luminescens"
                     /strain="W14"
                     /db_xref="taxon:29488"
                     /clone="PLG01100"
                     /clone_lib="Photorhabdus luminescens strain W14 M13
                     library"
                     /dev_stage="primary phase variant"
                     /note="Genomic DNA from strain W14 was size selected (1-2
                     kb) and then cloned into M13 Janus."
BASE COUNT            181 a 159 c 131 g 161 t 4 others
ORIGIN
alignment_scores:
Quality: 190.00      Length: 118
Ratio: 2.235        Gaps: 2
Percent Similarity: 72.034 Percent Identity: 36.441
```

```
alignment_block:
US-09-701-711-2 x AQ990340/rev ..

Align seg 1/1 to reverse of: AQ990340 from: 1 to: 636

8 GlyPheGlnValSerAlaMetThrMetAlaValMetValMet..... 22
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
356 GGAAGACGCATTACACGATGGCGATGAAAGATTACTCATAGCGTGCT 307
23 .....SerThrHisAlaGlnAlaAlaAsp...PheMetAlaAsnA 35
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
306 GCTGTTCGGCAGCCCACTGCATACGGTGCAGACGGATTCTAGTTCAGG 257
35 spileThrIleThrGluGlnArgValThrIleGluSerLeuGlnSer 51
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
256 ACATTCACTTTGAAGTCTTCAACGCGTCGCGTAGGTGCGAGCATTTATG 207
52 ValLeuProPheArgLeuGlyGlnValValSerGluAsnGlnLeuAlaAs 68
: : : : : : : : : : : : : : : : : : : : : : : : : :
206 AATATGCCAGTTCCGCTAGGTGATACGCTCAGCGATGAAGATATCGGTG 157
68 pGlyValLysAlaLeuTyrAlaThrGlyAsnPheSerAspValGlnValT 85
: : : : : : : : : : : : : : : : : : : : : : : : : :
156 TACCATTTCAGCACTATTCTCTACTGCTGTAATTGAAGACCTTCGTGTCC 107
85 yrHisGlnGluGlyArgIleIleTyrGlnValThrGluArgProLeuIle 101
: : : : : : : : : : : : : : : : : : : : : : : : : :
106 TCGGTGATGCAATACACTATTCTTCAGGTAAAGAACGCCGACTATT 57
102 AlaGluIleAsnPheGluGlyAsnArgLeuIleProLysGluGlyLeuGl 118
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
56 GCCAGCATCACTTTCTCCGCAATAATTCGTTGAAGATGACATGTTCAA 7
118 nGlu 119
: : : :
6 ACA 3

seq_name: gb_gss:AQ990755

seq_documentation_block:
LOCUS AQ990755 525 bp DNA linear GSS 14-AUG-2000
DEFINITION Rfc01576 photorhabdus luminescens strain W14 M13 library
Photorhabdus luminescens genomic clone PLG01576, DNA sequence.
ACCESSION AQ990755
VERSION AQ990755.1 GI:9649349
KEYWORDS GSS.
SOURCE Photorhabdus luminescens.
ORGANISM Photorhabdus luminescens
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Photorhabdus.
REFERENCE 1 (bases 1 to 525)
AUTHORS ffrench-Constant R.H., Waterfield,N., Burland,V., Perna,N.T.,
Daborn,P.J., Bowen,D. and Blattner,F.R.
TITLE A genomic sample sequence of the entomopathogenic bacterium
Photorhabdus luminescens W14: potential implications for virulence
Appl. Environ. Microbiol. 66 (8), 3310-3329 (2000)
JOURNAL 20378633
MEDLINE
COMMENT Contact: ffrench-Constant RH
Department of Biology and Biochemistry
University of Bath
South Building, Bath BA2 7AY, UK
Tel: (44) 1225 826621
Fax: (44) 1225 826779
Email: bsr@bath.ac.uk
This is one of 2,122 random reads from the M13 library. For
annotation of identified clones (BLASTX, BLASTN and mapping to E.
coli K12 genome) please see ffrench-Constant et al. 2000, Nucleic
Acids Res.
Seq primer: M13 Forward
Class: shotgun.
Location/Qualifiers
1..525
/organism="Photorhabdus luminescens"

/strain="W14"
/db_xref="taxon:29488"
/clone="PLG01576"
/clone_lib="Photorhabdus luminescens strain W14 M13
library"
/dev_stage="primary phase variant"
/notes="Genomic DNA from strain W14 was size selected (1-2
kb) and then cloned into M13 Janus."
BASE COUNT 160 a 94 c 114 g 154 t 3 others
ORIGIN

alignment_scores:
Quality: 182.50 Length: 178
Ratio: 1.534 Gaps: 4
Percent Similarity: 66.854 Percent Identity: 29.775

alignment_block:
US-09-701-711-2 x AQ990755 ..

Align seg 1/1 to: AQ990755 from: 1 to: 525

196 SerAspAlaAspLeuIleAspValLeuAlaIleLysAspAsnLysIleAs 212
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
2 TCTTCTGATGAGTTATTGAATCGTTTCCAACTTAGAGATGATGTC..... 46
212 nPro.....LeuSerLysAlaAspArgTyrThrGlnGluLysLeuV 226
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
47 .CCGTGGTGGAAATTGACTGCCGATCAGAAATATCAGAAACAAAACCTGA 95
226 alThrSerLeuGluAsnLeuArgAlaTyrTyrLeuAsnAlaGlyPheVal 242
: : : : : : : : : : : : : : : : : : : : : : : : : :
96 CTGCTGACCTTGAAGCATTCGCAGATTTTACCTTGATCGCGGTTACGCC 145
243 ArgPheGluIleLysAspAlaLysLeuAsnIleAsnGluAspLysAsnAr 259
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
146 CGTTTAAACATTTGATTCGACTCAGTCTGTTGACGCCAGATAAAAAAGG 195
259 gilePheValGluIleSerLeuHisGluGlyGluGlnTyrArgPheGlyG 276
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
196 TATTATGTCAGTAAATATACTGAAGGGATCAATACAAAATATATCAG 245
276 lnThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAlaGluLeuGluAla 292
: : : : : : : : : : : : : : : : : : : : : : : : : :
246 GTATTGACCTGACCGGTAATATGCGGGTTATCAGTCAGAAATTTACTAA 295
293 LeuLeuLysPheLysAlaGluGlyPheSerGlnAlaMetLeuGluGl 309
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
296 CTGGCTGTCAITTGACCTGGATCTCTGTATAACGGGACTCAGGTGACTAA 345
309 nThrThrAsnAsnIleSerThrLys.PheGlyAspAspGlyTyrTyrTyr 325
: : : : : : : : : : : : : : : : : : : : : : : : : :
346 AATGGAAATGACATCAAGAAATTTGCTTGGTCGCTATGTTATCTCTAC 395
326 AlaGlnIleArgProValThrArgIleAsnAspGluSerArgThrValAs 342
: : : : : : : : : : : : : : : : : : : : : : : : : :
396 CCACGAGTGAT.GACGACCTGAATCAATCAATGATCAAGACAGACAGTAA 444
342 pValGluTyrTyrIleAspProValHisProValTyrValArgArgIleA 359
: : : : : : : : : : : : : : : : : : : : : : : : : :
445 ACTGCATGTCAATATTTGATCGNGTNAACCGTTCTTCTATGCTGCTAAATTC 494
359 snPhe.ThrGlyAsnPheLysThrGlnAsp 368
: : : : : : : : : : : : : : : : : : : : : : : : : :
495 GTTTCCTTCTGTTGTAATGACACCACCTAAAGAT 524

seq_name: gb_est1:AW947726

seq_documentation_block:
LOCUS AW947726 278 bp mRNA linear EST 31-MAY-2000
DEFINITION RC0-MT0004-130300-011-c06 MT0004 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW947726
VERSION AW947726.1 GI:8125409
```

```

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.K., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL
MEDLINE
COMMENT
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-RCO-MT0004-130
300-011-c064t3=2000-03-13&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 61
High quality sequence stop: 278.
FEATURES
source
1. .278
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MT0004"
/dev_stage="Adult"
/note="Organ: marrow; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT
67 a 82 c 74 g 55 t
ORIGIN
alignment_scores
Quality: 179.00 Length: 90
Ratio: 2.486 Gaps: 4
Percent Similarity: 80.000 Percent Identity: 48.889
alignment_block
US-09-701-711-2 x AW947726 ..
Align seg 1/1 to: AW947726 from: 1 to: 278
416 AspGlnValAspValAsnPhValGluGluGlnProSerGlySerSe 432
|||||
17 GACCAGGTAGATGTGAACCTAGCGCGTAGAAGAACAAGCCTCAGGTTCCAT 66
|||||
432 rThrIleAlaAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheA 449
|||||
67 CACGCCAGCGTCGGTAGCGTAGAGATGCGGTTGATCCTGGGTGGTT 116
|||||
449 spValSerGlnAsnAsnPhMetGly.ThrGlyLysHis.ValAsnAlaAs 465
|||||
117 CCATCAGCAGACAACACTCTCGGATACCGGTAAACCGATGTGTCATCG 166
|||||
465 erPheSerArgSerGlu.ThrArgGluValTyrSerLeuGlyMetThrAs 481
|||||
167 GCCTGACCCGAGCGGAATPACGAGACCGACCAACTTCGCGTATACCGA 216
|||||
481 nProTyrPheThrValAsnGlyValSerGlnSerLeu.SerGlyTyrTyr 497
|||||
217 CCCCCTACTGGAGCTGACAGCGGTGAGCGTGGTGCTACTACGCCCTTCTAC 266
|||||
498 ArgLysThrLys 501
|||||
267 CGCACCACCGAA 278
|||||
seq_name: gb_gss:BH371368
seq_documentation_block:
LOCUS BH371368 776 bp DNA linear GSS 10-DEC-2001
DEFINITION AG-ND-175A22.TR ND-TAM Anopheles gambiae genomic clone AG-ND-175A22
, DNA sequence.
ACCESSION BH371368
VERSION BH371368.1 GI:17317493
KEYWORDS GSS.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea
; Anopheles.
REFERENCE
1 (bases 1 to 776)
AUTHORS Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
TITLE Direct Submission of BAC-end sequences from Anopheles gambiae
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: AG-ND-175A22.TF
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research
(FIGR). The BAC library was generated from A. gambiae PEST strain
DNA. All DNA was extracted from newly hatched first instar larvae
to minimize the inclusion of DNA from microorganisms that inhabit
the gut. The DNA is derived from mixed sexes of larvae. The BAC
library was constructed at Texas A&M University BAC Center
University, College Station, Texas 77843-2123, USA using a HindIII
partial digest.
Seq primer: M13 Rev
Class: BAC ends.
FEATURES
source
1. .776
Location/Qualifiers
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone_lib="AG-ND-175A22"
/note="Vector: pECBAC1; Site_1: HindIII"
BASE COUNT
213 a 140 c 120 g 303 t
ORIGIN
alignment_scores
Quality: 150.50 Length: 240
Ratio: 1.060 Gaps: 9
Percent Similarity: 59.167 Percent Identity: 22.500
alignment_block
US-09-701-711-2 x BH371368/rev ..
Align seg 1/1 to reverse of: BH371368 from: 1 to: 776
65 GlnLeuAlaAspGlyValLysAlaLeuTyrAlaThrGlyAsnPhSerAs 81
|||||
698 AGACTGAGTACAGCTATCAAAAGCTTTGGGATAGCCCAATATTTCTGA 649
|||||
81 pValGlnValTyr.....HisGlnGluGlyArgIleI 92
|||||

```



```

|||||:|||||
648 AGTTGAAGTATATGTCAGACATTATAGCACACATTAAAGGCACAAATTTG 599
92 leTyrGlnValThrGlu.....ArgProLeuIleAlaGluIleAsn 105
598 TCTTAATCTCTCTTTCAGGATTTAAAGAACTTGGTAAGTAAATTTCC 549
106 PheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluGlyLeu 122
548 GGAAGAAGCA.....ATCAAAATATCTAAAGAACGAAAACTTTATTAA 508
122 sAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrValGlnM 139
507 GGATAATAACCTGAAGCCTGTGTATGAATAATACCGAGAACCCTGGTTACTA 458
139 etIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyrTyrAsn 155
457 ACCTAAAGCATATGTACTCTCAGCAATACATCAAAAGGATTCCTCCGAT 408
156 ThrGluIleThrVal.....LysGlnThrMetLeuAspGlyAsnArgVa 170
407 GCTAAGATCACCATTGAAGATATAAATCAATGCAAAAGGATCTAGCTTAT 358
170 lLysLeuAspMetThrPheAlaGluGlyLysProAlaArgValValAspI 187
357 CGACTGGACCATTAATGTAGATAAAGGAAAAACGTTTAAAGATAGACCGTA 308
187 leAsnIleIleGlyAsnGlnHisPheSerAspAlaAspLeu.Ile...As 202
307 TAGATTTGAAGGAAACAAACAGTGTCTTCTCACTAAACTTCGTAAAC 258
202 pValLeuAlaIleLysAspAsnLys.....IleAsnProLeuSerLysA 217
257 GGTTTTAAATAACGAAACAAAGAGATTCCTTCTGTTGCTTAAGC 208
217 laAspArgTyrThrGlnGluLysLeuValThrSerLeuGluAsnLeuArg 233
207 CTTCGAATTTATCAAGATAAATATGAAGAGATAAGAGACTCTTGTA 158
234 AlalysTyrLeuAsnAlaGlyPheValArgPheGluIleLysAspAlaLy 250
157 GACTACTACAACTCTTTAGGATTT.....AGAGATATAGAG 123
250 sLeu.....AsnIleAsnGluAspLysAsnArgIlePheValGluI 264
122 AGTAGTGCAGACTCTGTTTCGCGTAATACTAAAGGATACACATCAAG 73
264 leSerLeuHisGluGlyGluGlnTyrArgPheGlyGlnThrGlnPheLeu 280
72 TAAAGTAGATAGAGGGTAAAAAATACTATATCGGAGATATTACATTCGTA 23
281 GlyAsnLeuThrTyrThr 286
22 GGAATACAGATTTTCT 5

```

seq\_name: gb\_est2:BI903558

```

seq_documentation_block:
LOCUS      BI903558                      767 bp    mRNA    linear    EST 16-OCT-2001
DEFINITION 603168388F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5256452 5',
            mRNA sequence.
ACCESSION  BI903558
VERSION    BI903558.1 GI:16165506
KEYWORDS  EST.
SOURCE     house mouse.
ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 767)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.

```

Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA sequencing by: Incyte Genomics, Inc.  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: L1AM11647 row: p column: 21  
High quality sequence start: 22  
High quality sequence stop: 764.  
Location/Qualifiers  
source  
1. .767  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5256452"  
/tissue\_type="pooled lung tumors"  
/lab\_host="DH10B (phage-resistant)"  
/notes="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: NotI; Site\_2: EcoRI; 1st strand cDNA was prepared from mRNA obtained from pooled lung tumors with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGCGCGCTCTGTTTTTTTTTTTTTTT 3']  
Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73D vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

## FEATURES

## source

BASE COUNT 175 a 199 c 221 g 171 t 1 others  
ORIGIN  
alignment\_scores:  
Quality: 139.50 Length: 239  
Ratio: 1.107 Gaps: 10  
Percent Similarity: 52.720 Percent Identity: 25.105  
alignment\_block:  
US-09-701-711-2 x BI903558 ..  
Align seg 1/1 to: BI903558 from: 1 to: 767  
590 SerLeuAspArgProValPheProThrGlnGlyMetSerHisSerValas 606  
3 ACTCTCAACCCCTGCTGTATTGCCAAGAAGAGGGGCTTGTTCAAAGTCAA 52  
606 pleuThrVal...GlyPhe.....GlyAspLysThrHisGlnLysValV 620  
53 CCAGGAGCTGGCAGGCTACACTGGAGAGATGTGAGCTTCATCAAGGAAG 102  
620 alTyrGlnGlyAsnIleTyrArgProPheIleLysLysSerValLeuArg 636  
103 ACTTTGAGCTTCAGCTGAATAAGCGCTCGCTTGAGCTCGGTATTCTCC 152  
637 GlyTyrAlaLysLeuGlyTyrGlyAsnLeuPro..... 648  
153 .....ACGTCTCTGGGTGGATGCTGGTGGCCATCGGTGACAAAGCC 196  
649 .....PheTyrGluAsnPheTyrAlaGlyTyrGlySerValArg 663  
197 ATCCAGCATTCGTGACAGGTTTACTGGAGGCCGCCAGAGTGTCCGAG 246  
663 lyTyrAspGlnSerSerLeuGlyProArgSerGlnAlaTyrLeuThrAla 679  
247 GATTAGCATGCACAGCATTTGAGCCCAAGAGTGA..... 281  
680 ArgArgGlyGlnGlnThrThrLeuGlyGluValValGlyGlyAsnAlaLe 696  
282 .....GGAGATTACCTGGCGCGGCGGCGCTTA 707



seq\_name: gb\_est2:BE678332

seq\_documentation\_block:

LOCUS BE678332 547 bp mRNA linear EST 18-APR-2001  
 DEFINITION df54c02.x1 xenopus laevis unfertilized egg cDNA library Xenopus  
 laevis cDNA clone IMAGE:3743138 3' similar to SW:YDF6\_SCHPO Q10478  
 HYPOTHETICAL 51.8 KD PROTEIN C17C9.06 IN CHROMOSOME I.; mRNA  
 sequence.

ACCESSION BE678332  
 VERSION BE678332.1 GI:10059901

KEYWORDS EST.

SOURCE African clawed frog.

ORGANISM Xenopus laevis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;  
 Xenopodinae; Xenopus.

1 (bases 1 to 547)

REFERENCE  
 AUTHORS Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.,  
 Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person  
 , B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,  
 Waterston,R. and Wilson,R.

WashU Xenopus EST project, 1999

Unpublished (1999)

Contact: Sandy Clifton, Ph.D.

WashU Xenopus EST project, 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Library constructed by Bruce Blumberg

Library normalized by Jihwan Song

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: xenopus clone distribution information for

this library can be found through Research Genetics, visit their

web page at: <http://www.resgen.com/> Please reference the id listed

below when ordering this clone: Source lab clone id - xlnneg002b03

Seq primer: -40UP from Gibco

High quality sequence stop: 477.

#### FEATURES

source

1..547

/organism="Xenopus laevis"

/db\_xref="taxon:8355"

/clone\_image="3743138"

/clone\_lib="Xenopus laevis unfertilized egg cDNA library"

/tissue\_type="unfertilized egg"

/lab\_host="Top-10 F"

/note="Vector: pBluescript SK-; Site1: EcoRI; Site2:

XhoI; cDNA was prepared from 2ug of poly A+ RNA.

EcoRI-XhoI cut cDNA was then ligated into Unizap-XR

(Stratagene) with EcoRI at the 5' end and XhoI at the 3'

end. SS-library phagemids were prepared by mass excision

from the original library and normalized by hybridization

to biotinylated driver (prepared from the same library by

PCR) to Cot-omega of 11. After removal of hybrids and

excess driver by streptavidin sepharose chromatography,

the ss-phagemids were made double stranded and

electroporated into Top-10 F'. Original library

contruction by Bruce Blumberg (Blumberg et al., 1991

Science 253, 194-196; Hawlet et al., 1995, Genes Dev. 9,

2923-2935). Note: This is a Xenopus Gene Collection (XGC)

library."

BASE COUNT 131 a 161 c 136 g 119 t

ORIGIN

alignment\_scores:

Quality: 136.50

Ratio: 1.517

Percent Similarity: 53.571

Percent Identity: 27.381

alignment\_block:

US-09-701-711-2 x BE678332/rev ..

Align seg 1/1 to reverse of: BE678332 from: 1 to: 547

```

651 GluAsnPhetYrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667
      :::::|||||  |||||  |||||  |||||  |||||  |||||  :::::
544 GACCGGTTTTACCTGGGGGACCCACGAGCGTGGCAGGATTACGATGTA 495
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgGlyGlnG 684
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
494 CAGCACTGGCCCCAGACGAA..... 473

684 lnThrLeuGlyGluValValGlyGlyAsnAlaLeuAlaThrPheGly 700
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
472 .....GGCGACTATTGGCGCGAGAAGCTTACTGGGCTGGGGC 434
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
701 SerGluLeuLeuLeuLeuProPhe.....LysGlyAspTrpIl 714
      :::::  |||||  |||||  |||||  |||||  |||||  :::::
433 GTGCACTTATACACCCCTCTACCCCTTCGCGCGGACGCGGGGATTGG 384
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
714 eAspGlnValArgProValIlePheIleGluGlyGlnValPheAspT 731
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
383 GGACCTTTTCAGACCCACTTCTCTCAACCTGGGAACCTC..... 341
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
731 hrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPheLysAspPro 747
      :::::  |||||  |||||  |||||  |||||  |||||  :::::
340 .....TGCAACCTGAACCTACGCGGGGAGGCCCC 314
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
748 GlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuTh 764
      :::::  |||||  |||||  |||||  |||||  |||||  :::::
313 AGGCCCATCTCGACGCGATTGGCTGAATGT..... 284
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
764 rGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrTrp...T 780
      :::::  |||||  |||||  |||||  |||||  |||||  :::::
283 .....ATCCGCTGGTCTCATGGGCGGGGCTAGTTCGCGCT 247
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
780 yrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeuAsnLys 796
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
246 TGGGGAATATCGCAGACTGGGAGCTGAATTATTGTATCCCTATGGGGT 197
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
797 LysGlnAsnAspGln...ThrAspThrValGlnPheGlnIleGlySerVa 812
      :::::  |||||  |||||  |||||  |||||  |||||  :::::
196 CAGAGTGGAGAGAGGATCTGCGATGGGTTTCAGTTCGCGCTGGAATCCG 147
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
812 lphe 813
      |||||
146 ATTC 143

```

seq\_name: gb\_est2:BE678801

seq\_documentation\_block:

LOCUS BE678801

DEFINITION df76g02.x1 Xenopus laevis oocyte non normalized Xenopus laevis cDNA  
 clone IMAGE:3745298 3' similar to SW:YNC6\_YEAST P53969 HYPOTHETICAL  
 54.4 KD PROTEIN IN KTR5-UME3 INTERGENIC REGION.; mRNA sequence.

ACCESSION BE678801

VERSION BE678801.1 GI:10060840

KEYWORDS EST.

SOURCE African clawed frog.

ORGANISM Xenopus laevis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;

Xenopodinae; Xenopus.

1 (bases 1 to 553)

REFERENCE Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.,

, Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person

, B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,

Waterston,R. and Wilson,R.

WashU Xenopus EST project, 1999

Unpublished (1999)

Contact: Sandy Clifton, Ph.D.

WashU Xenopus EST project, 1999

Washington University School of Medicine





|||||  
252 GTCCGAA 259

seq\_name: gb\_est2:BM119288

seq\_documentation\_block: 697 bp mRNA linear EST 01-FEB-2002  
LOCUS BM119288  
DEFINITION L0923D10-3 NTA Mouse Newborn Kidney cDNA Library (Long) Mus musculus cDNA clone L0923D10 3', mRNA sequence.

ACCESSION BM119288  
VERSION BM119288.1 GI:17087314

KEYWORDS EST.  
SOURCE house mouse.

ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 697)  
Piao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Pantano, S., Lim, M.K. and Ko, M.S.H.

TITLE Systematic Analyses of NIA Mouse Newborn Kidney cDNA Library

JOURNAL Unpublished (2001)  
COMMENT Contact: Dawood B. Dudekula  
Laboratory of Genetics  
National Institute on Aging/National Institutes of Health  
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
Email: cdnaeigsun.grc.nia.nih.gov  
Plate: L0923 row: D column: 10  
Seq primer: -2LM13 Forward  
High quality sequence stop: 697  
POLYA-res.

FEATURES

Location/Qualifiers  
1..697  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="niaEST:L0923D10-3"  
/db\_xref="taxon:10090"  
/clone="L0923D10"  
/clone\_lib="NIA Mouse Newborn Kidney cDNA Library (Long)"  
/tissue\_type="Newborn Kidney"  
/dev\_stage="Newborn"  
/lab\_host="DH10B"  
/note="Organ: kidney; Vector: pSPORT1 (Invitrogen); Site\_1: SalI; Site\_2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11:1553-1558 (2001). [PMID:11544199]). In brief ,double-stranded cDNAs were synthesized with an oligo(dT) primer [Invitrogen: 5'- pGACTAGTCTAGATCGAGCGCGCCCTTTT-3'] from 26 microgram of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker LL-Sal4, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes, and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with ligation mixture by the standard chemical method. The average insert size is about 3.0 kb. The library was constructed by Yulan Piao (NIA)."

BASE COUNT 159 a 207 c 185 g 146 t  
ORIGIN

alignment\_scores:  
Quality: 134.00 Length: 195  
Ratio: 1.354 Gaps: 8  
Percent Similarity: 50.769 Percent Identity: 26.154

alignment\_block:

US-09-701-711-2 x BM119288/rev ..

Align seg 1/1 to reverse of: BM119288 from: 1 to: 697

639 AlaLysLeuGlyTyrGly.....AsnAsnLe 647  
|||||  
621 GTCGCCTTGGAGTCGGTATTCTCCACGCTCTCTGGGTGGAATGCTGT 572  
647 uPro.....PheTyrGluAsnPheTyrAlaGlyG 657  
|||||  
571 GCCATCGGTGACAGCCATTCAGCATTCGTGACAGGTTTACCTGGGAG 522  
657 lYtYrGlySerValArgGlyTyrAspGlnSerSerLeuGlyProArgSer 673  
|||||  
521 GCCCCAGAGTGTCCGAGATTAGCATGCACAGCATTCGACCCAGAGT 472  
674 GlnAlaTyrLeuThrAlaArgArgGlyGlnGlnThrThrLeuGlyGluVa 690  
|||||  
471 GAA.....GGAGATTAA 461  
690 lValGlyGlyAsnAlaLeuAlaThrPheGlySerGluLeuLeuLeuProL 707  
|||||  
460 CTTGGCGCGGAGGCTACTGGGCTGGGGCTGCACCTCTACACCCAC 411  
707 euProPheLys.....GlyAspTyrPheAspGlnValArgProVal 720  
|||||  
410 TGCCTTCGGCCAGCCAGGTGGCTTCGGAGAGCTTTTCAGAACTCAC 361  
721 lIePheIleGluGlyGlnValPheAspThrThrGlyMetAspLysGl 737  
|||||  
360 TTTTCTCAATCGGGCAACCTG..... 337  
737 nThrIleAspLeuThrGlnPheLysAspProGlnAlaThrAlaGluGlnA 754  
|||||  
336 ...TCCAACCTCAACTATGCTGAGGGCCCCAAAGCCCATATCCGGAAGC 291  
754 snAlaLysAlaAlaAsnArgProLeuLeuThrGlnAspLysGlnLeuArg 770  
|||||  
290 TAGCTGAGTGC.....ATCCGC 274  
771 TyrSerAlaGlyValGlyAlaThrTrp...TyrThrProIleGlyProle 786  
|||||  
273 TGTCTCTATGAGCAGCGCTCGCTCCGACTTGCCACATCGCTCGGT 224  
786 uSerIleSerTyrAlaLysProLeuAsnLysLysGlnAsnAspGln...T 802  
|||||  
223 GGAGCTGAACTACTGCAATTCCTATGGCGCTGAGGGGGCGGACAGGATT 174  
802 hrAspThrValGlnPheGlnIleGlySerValPhe 813  
|||||  
173 GTGATGGTGTCCAGTTTGGAGCTGGGATTCGGTTC 139

seq\_name: gb\_est2:BG303393

seq\_documentation\_block:

LOCUS BG303393 675 bp mRNA linear EST 23-FEB-2001  
DEFINITION f155a04.x1 Sugano Kawakami zebrafish DNA Danio rerio cDNA clone 3817615 3' similar to TR:Q9Y512 Q9Y512 DJ796117.2 ; mRNA sequence.

ACCESSION BG303393  
VERSION BG303393.1 GI:13100920

KEYWORDS EST.  
SOURCE zebrafish.

ORGANISM Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes ; Cyprinidae; Danio.  
1 (bases 1 to 675)

REFERENCE  
AUTHORS

Sugano, S., Kawakami, K., Johnson, S., Li, F., Marra, M., Eddy, S., Hillier, L., Clifton, S., Allen, M., Gibbons, M., Jost, S., Kucaba, T., Martin, J., Pape, D., Steptoe, M., Underwood, K., Theising, B., Ritter, E., Bowers, Y., Wylie, T., Waterston, R. and Wilson, R.  
WashU Zebrafish EST Project 1999

TITLE

JOURNAL  
COMMENT

Unpublished (1999)  
 Contact: S.L. Johnson  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA  
 Sequencing by: Washington University Genome Sequencing Center Clone  
 distribution information can be found through the I.M.A.G.E.  
 Consortium/LLNL, send email to: info@image.llnl.gov  
 Seq primer: T7 from Gibco  
 High quality sequence stop: 476.

## FEATURES

## source

```
1. .675
/organism="Danio rerio"
/strain="AB"
/db_xref="taxon:7955"
/clone_lib="Sugano Kawakami zebrafish DRA"
/sex="mixed (one male and one female, including
unfertilized eggs)"
/dev_stage="adult"
/lab_host="DH10B (phage resistant)"
/note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG);
Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
with an oligo(dT) primer [ATGTGGCTTTTTTTTTT];
double-stranded cDNA was ligated to a DraIII adaptor
[TTGTGGCCCTACTGG], digested and cloned into distinct DraIII
sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
CACCATGTG). XhoI should be used to isolate the cDNA
insert. Size selection was performed to exclude fragments
<1.5kb. Library constructed by Dr. Sumio Sugano
(University of Tokyo Institute of Medical Science) and
kindly donated by Dr. Koichi Kawakami. Custom primers for
sequencing: 5' end primer CTTCGTCTAAAGCTCGG and 3' end
primer CGACCTCGAGCTCGAGACA."
```

BASE COUNT  
 ORIGIN

## alignment\_scores:

Quality: 133.50 Length: 168  
 Ratio: 1.483 Gaps: 6  
 Percent Similarity: 53.571 Percent Identity: 27.381

## alignment\_block:

US-09-701-711-2 x BG303393/rev ..

Align seg 1/1 to reverse of: BG303393 from: 1 to: 675

```
651 GluAspPheTyrAlaGlyTyrGlySerValArgGlyTyrAspGlnSe 667
      :::::||||| ||||| ||||| ||||| ||||| |||||
642 GACAGGTTTATTGGTGGTGCTCTACCAGTGTGAGGGCTTTAGCATGA 593
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgGlyGlnG 684
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
592 CAGCATCGGCCCTCAGAGTGAA..... 571
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
684 lnThrThrLeuGlyGluValValGlyGlyAsnAlaLeuAlaThrPheGly 700
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
570 .....GGAGATTACCTCGCGCGGAGAGCGCTATTGGCGTGGGGA 532
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
701 SerGluLeuIleLeuProLeuProPhe.....LysGlyAspTrpI 714
      :::::||||| ||||| ||||| ||||| ||||| |||||
531 GTTCACCTACACGCTCTTCATTACGCGCTGGCGAGGTGATTCGG 482
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
714 eAspGlnValArgProValIlePheIleGluGlyGlnValPheAspT 731
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
481 AGATCTTTTCAGACACATTTTCTCAATGCCGGAACCTCG..... 439
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
731 hrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPheLysAspPro 747
      :::::||||| ||||| ||||| ||||| ||||| |||||
```

```
438 .....TGTAACCTCAACTACGGGAAGGCCG 412
748 GlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuTh 764
      :::::||||| ::::: ||||| :::::
411 CGGGCTACCTCAGCAAACTGGCCGAGTGC..... 382
764 rGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrTrp...T 780
      :::::||||| ||||| ||||| ||||| |||||
381 .....ATCCGCTGGTCGTATGCGAGCAGGTATTGTGCTGGGTC 345
780 yrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeuAsnLys 796
      ||||| ||||| ::::: ||||| ||||| |||||
344 TGGGGAACATCGCTGACGTGGAGCTAAATCTCATTCCTATGGGGTC 295
797 LysGlnAsnAspGln...ThrAspThrValGlnPheGlnIleGlySerVa 812
      :::::||||| ||||| ||||| ||||| |||||
294 CAGAGCGGACAGGATATGTGACGGAGTGCAGTTTGGGGCCGGAATCCG 245
812 lPhe 813
      |||
244 CTTC 241
seq_name: gb_est2:BG344329
seq_documentation_block:
LOCUS BG344329 804 bp mRNA linear EST 22-OCT-2001
DEFINITION HVSMEG0008N21f Hordeum vulgare pre-anthesis spike EST library
HVCDNA0008 (white to yellow anther) Hordeum vulgare cDNA clone
HVSMEG0008N21f, mRNA sequence.
ACCESSION BG344329
VERSION BG344329.2 GI:16317704
KEYWORDS EST.
SOURCE barley.
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 804)
AUTHORS Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton
,R.D., Close,S.J., Oates,R. and Main,D.
Development of a genetically and physically anchored EST resource
for barley genomics: Morex pre-anthesis spike cDNA library
Unpublished (2001)
On Feb 27, 2001 this sequence version replaced gi:13156658.
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hq bases = 535
Seq primer: AATTACCCCTCACTAAAGG
High quality sequence stop: 656.
Location/Qualifiers
1. .804
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMEG0008N21f"
/cclone_lib="Hordeum vulgare pre-anthesis spike EST library"
HVCNA0008 (white to yellow anther)"
/tissue_type="pre-anthesis spike"
/lab_host="SOLR"
/note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI;
Plants were grown in the greenhouse at the University of
California, Riverside (Fenton, SJ Close, TJ Close). Whole
spike with awns trimmed were collected at white, green and
yellow anther stages (Fenton). Total RNA was prepared from
each pool, equal quantities of all three RNA pools were
combined, poly(A) RNA was purified from the mixture, one
primary unamplified cDNA library was made, and 1 million
```





polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-c20 library is a non-normalized library constructed from rat penis tissue. For a detailed description of the library from which this clone was derived, please visit our web site at [ratest.eng.uiowa.edu](http://ratest.eng.uiowa.edu). The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)

TAG\_LIB=UI-R-C20  
TAG\_TISSUE=penis  
TAG\_SEQ=TTGCGGAACA"

BASE COUNT 126 a 164 c 153 g 130 t  
ORIGIN

alignment\_scores:  
Quality: 132.50 Length: 168  
Ratio: 1.472 Gaps: 6  
Percent Similarity: 53.571 Percent Identity: 26.786

alignment\_block:  
US-09-701-711-2 x BI278122/rev ..

Align seg 1/1 to reverse of: BI278122 from: 1 to: 573

```

651 GluAsnPheTyrAlaGlyClyTyrGlySerValArgGlyTyrAspGluSe 667
      ::::::::::::::::::::|
558 GACAGGTTTACCTGGAGGCCACGAGCGTCGAGGATTAGCATGCA 509
      ::::::::::::::::::::|
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
      ::::::::::::::::::::|
598 CAGCATCGACGCCAGAGCGAA..... 487
      ::::::::::::::::::::|
684 InThrLeuGlyGluValGlyGlyAsnAlaLeuAlaThrPheGly 700
      ::::::::::::::::::::|
486 .....GGAGATTACCTGGCGGGAGGCATACCTGGCCCGGGGC 448
      ::::::::::::::::::::|
701 SerGluLeuLeuLeuProLeuProPheLys.....GlyAspTrpI 714
      ::::::::::::::::::::|
447 CTGCACCTCTATACCCACCTGCTTTCCGGCCAGGCGGTGCTTCGG 398
      ::::::::::::::::::::|
714 eAspGlnValArgProValIlePheIleGlyGlyGlnValPheAspT 731
      ::::::::::::::::::::|
397 AGAGCTTTTCAGACTCACTTTTCTCAATGCTGCGCAACCTG..... 355
      ::::::::::::::::::::|
731 hrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPheLysAspPro 747
      ::::::::::::::::::::|
354 .....TGCACCTCACTAGCTGAGGCGGCC 328
      ::::::::::::::::::::|
748 GlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuTh 764
      ::::::::::::::::::::|
327 AGAGCCACATCCGGAAGCTGCTCAATGC..... 298
      ::::::::::::::::::::|
764 rGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrTrp...T 780
      ::::::::::::::::::::|
297 .....ATCCGCTGGTCTATGTTGTCAGGCGATCGTCCTCCGCG 261
      ::::::::::::::::::::|
780 yrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeuAsnLys 796
      ::::::::::::::::::::|
260 TTGGCAACATTGCTCGGTGGAGTGAACACTACTGCATTCTCATGGGTG 211
      ::::::::::::::::::::|
797 LysGlnAsnAspGln...ThrAspThrValGlnPheGlnIleGlySerVa 812
      ::::::::::::::::::::|
210 CAGCGAGCGCAGGATTGTGTGGCTCCAGTTTGGAGCTGGGATTCCG 161
      ::::::::::::::::::::|
812 lPhe 813
      ::::::::::::::::::::|
160 GTTC 157

```

seq\_name: gb\_est2:BJ054028

seq\_documentation\_block:

LOCUS BJ054028 721 bp mRNA linear EST 11-DEC-2001  
DEFINITION BJ054028 NIBB Mochii normalized Xenopus neurula library

laevis cDNA clone XL049a15 3', mRNA sequence.

ACCESSION BJ054028  
VERSION BJ054028.1 GI:17499034

KEYWORDS EST  
SOURCE African clawed frog.

ORGANISM

Xenopus laevis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
Xenopodinae; Xenopus.

REFERENCE 1 (bases 1 to 721)

AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara  
Y.

TITLE Expressed genes in X. laevis embryo

JOURNAL Unpublished (2001)

COMMENT Contact: Tadasu Shin-i

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshini@genes.nig.ac.jp.

FEATURES

Location/Qualifiers

source

1..721

/organism="Xenopus laevis"

/db\_xref="taxon:8355"

/clone="XL049a15"

/clone\_lib="NIBB Mochii normalized Xenopus neurula

library"

/tissue\_type="whole embryo"

/dev\_stage="stage 15"

BASE COUNT 162 a 218 c 183 g 157 t 1 others

ORIGIN

alignment\_scores:

Quality: 132.00 Length: 214

Ratio: 1.211 Gaps: 8

Percent Similarity: 50.935 Percent Identity: 26.168

alignment\_block:

US-09-701-711-2 x BJ054028/rev ..

Align seg 1/1 to reverse of: BJ054028 from: 1 to: 721

```

612 GlyAspLysThrHisGlnLysValValTyrGlnGlyAsnIleTyrArgPr 628
      ::::::::::::::::::::|
687 GGGGACGTCAGTTTCTTAAGGAAGACTTTGAGTTCAGTTGAACAGGCA 638
      ::::::::::::::::::::|
628 oPheIleLysLysSerValLeuArgGlyTyrAlaLysLeuGlyTyrGlyA 645
      ::::::::::::::::::::|
637 GCTTACCTGGGACTCTGCTGTCG.....ACATCTCTGTGGGGAGAA 594
      ::::::::::::::::::::|
645 snAsnLeuPro.....PheTyrGluAsnPheTyr 554
      ::::::::::::::::::::|
593 TGTGTGTCGCGATCGGAGACAGCGCGACGAGTATCGACACCGGTTCTAC 544
      ::::::::::::::::::::|
655 AlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSerSerLeuGlyPr 671
      ::::::::::::::::::::|
543 CTGGGGGGACCCACGAGCGTGGAGGATTCAGCATGTACAGCATGGGCC 494
      ::::::::::::::::::::|
671 oArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnGlnThrThrLeuG 688
      ::::::::::::::::::::|
493 CCAGAGTGAA.....G 483
      ::::::::::::::::::::|
688 lyGluValValGlyGlyAsnAlaLeuAlaThrPheGlySerGluLeuIle 704
      ::::::::::::::::::::|
482 GTGACTATTGGCGGGGAAGCTTACTGGGCTGGGGCGTGCACATTATAC 433
      ::::::::::::::::::::|
705 LeuProLeuProPhe.....LysGlyAspTyrIleAspGlnValAr 718
      ::::::::::::::::::::|
432 ACCCCCTCCCTTCGCGCGGACGCGGGGATTCGGGAGCTTTTTCAG 383
      ::::::::::::::::::::|
718 gProValIlePheIleGlyGlyGlnValPheAspThrThrGlyMeta 735

```

```

382 GACTCACTTCTCTGTAACGCTGGGAACCTC..... 352
735 spLysGlnThrIleAsnProGlnPheLysAspProGlnAlaThrAla 751
351 .....TGCACACTTGAACATACGGGGAGGCGCCCGGGCCCATCTT 313
752 GluGlnAsnAlaLysAlaAsnArgProLeuLeuThrGlnAspLysG1 768
312 CGACGATTGGCTGAATGT..... 295
768 nLeuArgTyrSerAlaGlyValGlyAlaThrTrp...TyrThrProLeuG 784
294 .ATCCGCTGGTCCATATGCGGGCGGGCTAGTTCTGCGCTTGGGGAATATCG 246
784 lyProLeuSerIleSerTyrAlaLysProLeuAsnLysLysGlnAsnAsp 800
245 CGAGACTGGAACGAAATATTATGATCCCTATGCGGGTCCAGAGCGGAGAC 196
801 Gln...ThrAspThrValGlnPheGlnIleGlySerValPhe 813
195 AGGATCTCGGACGGGTCCAGTTCGGGCGCGGAATCCGATTC 154

```

seq\_name: gb\_est2:BJ088421

seq\_documentation\_block:  
LOCUS BJ088421 481 bp mRNA linear EST 12-DEC-2001  
DEFINITION BJ088421 N1BB Mochii normalized Xenopus tailbud library Xenopus laevis cDNA clone XL093a01 3', mRNA sequence.

ACCESSION BJ088421

VERSION BJ088421.1 GI:17585682

KEYWORDS EST.

SOURCE African clawed frog.

ORGANISM

Xenopus laevis  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;  
Xenopodinae; Xenopus.

REFERENCE 1 (bases 1 to 481)

AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara Y.

TITLE Expressed genes in X. laevis embryo

JOURNAL Unpublished (2001)

COMMENT Contact: Tadasu Shin-i

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshini@genes.nig.ac.jp.

FEATURES

Location/Qualifiers

1..481

/organism="Xenopus laevis"

/db\_xref="taxon:8355"

/clone="XL093a01"

/library="N1BB Mochii normalized Xenopus tailbud library"

/tissue\_type="whole embryo"

/dev\_stage="stage 25"

98 a 161 c 132 g 90 t

BASE COUNT

ORIGIN

alignment\_scores:

Quality: 131.50

Ratio: 1.478

Percent Similarity: 52.976

Length: 168

Gaps: 6

Percent Identity: 27.381

alignment\_block:

US-09-701-711-2 x BJ088421/rev ..

Align seg 1/1 to reverse of: BJ088421 from: 1 to: 481

651 GluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667

```

438 GACGGTCTACCTGGGGGACCCACGAGCTGCGGAGGATTCAGCATGTA 389
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
388 CAGCATTTGGCCCCCAGAGTGAA..... 367
684 lnThrThrLeuGlyGluValValGlyGlyAsnAlaLeuAlaThrPheGly 700
366 .....GGTGACTATTATTGGCGGGGAAGCTTACTGGGCTGGGGGC 328
701 SerGluLeuIleLeuProLeuProPhe.....LysGlyAspTrpI1 714
327 GTGCACATTATACACCCCTCCCTTCGGCGCGGACGCGGGGATTCGG 278
714 eAspGlnValArgProValIlePheIleGluGlyGlyGlnValPheAspT 731
277 GGACCTTTTTCAGGACTCATTCTTCTCAGCGCTGGGAACCTC..... 235
731 hrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPheLysAspPro 747
234 .....TGCAACTTGAACCTAGCGGGGAGGGCCCC 208
748 GlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuTh 764
207 GGGGCCCATCTTCGACGATTGGCTGAATGT..... 178
764 rGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrTrp...T 780
177 .....ATCCGCTGGTCCTATGGGCGGGGCTAGTTCTTCGCGCT 141
780 yrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeuAsnLys 796
140 TGGGGAATATCGGAGACTGGAAGTGAATATTGTATCCCTATGGGGGTC 91
797 LysGlnAsnAspGln...ThrAspThrValGlnPheGlnIleGlySerVa 812
90 CAGAGCGGAGACAGGATCTCGAGGGGTCCAGTTCGGCGCGCGGAATCCG 41
812 lPhe 813
40 ATTC 37

```

seq\_name: gb\_est2:BM233570

seq\_documentation\_block:

LOCUS BM233570

DEFINITION K0338G02-3 N1A Mouse Osteoblast cDNA Library (Long) Mus musculus

CDNA clone K0338G02 3', mRNA sequence.

ACCESSION BM233570

VERSION BM233570.1 GI:17868840

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 580)

AUTHORS Piao,Y., Kargul,G.J., Dudekula,D.B., Qian,Y., Luo,A., Carter,M.G.,

Umezawa,A. and Ko,M.S.H.

Systematic Analyses of N1A Mouse Osteoblast cDNA Library (Long)

Unpublished (2001)

CONTACT: Dawood B. Dudekula

Laboratory of Genetics

National Institute on Aging/National Institutes of Health

333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA

Email: cdnae@igsun.grc.nia.nih.gov

Plate: K0338 row: G column: 02

Seq primer: -21M13 Forward

High quality sequence stop: 580

POLYA=Yes.

FEATURES

Location/Qualifiers

1..580

/organism="Mus musculus"

```

/strain="C3H/He mice"
/db_xref="niaEST:K0338G02-3"
/db_xref="taxon:10090"
/clone="K0338G02"
/clone_lib="NIA Mouse Osteoblast cDNA Library (Long)"
/tissue_type="Osteoblast"
/cell_line="KUS8-Al cells"
/lab_host="DH10B"
/Note="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2:
NotI; Mouse cDNA project by the Laboratory of Genetics,
National Institute on Aging (NIA), Intramural Research
Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is
a long-transcript enriched cDNA library (Ref. Genome Res.
11: 1553-1558 (2001)). [PMID: 115441199]. Total RNAs were
obtained from Dr. Akihiro Umezawa (Keio University School
of Medicine, Japan). Double-stranded cDNAs were
synthesized with an Oligo(dT) primer [Invitrogen:
5'-pGACTAGTCTAGATCGAGCGCGCCGCTTTTTTTTTT-3'] from
2.1 ug of total RNA, treated with T4 DNA polymerase, and
purified by ethanol-precipitation. The cDNAs were ligated
to lone-linker LL-SalI, purified by phenol/chloroform, and
separated from free linkers by Centricon 100. Then, the
cDNAs were amplified by long-range high fidelity PCR using
Ex Taq polymerase (Takara) with a primer Sal4-S. The
products were purified by phenol/chloroform and Centricon
100. The cDNAs were digested with SalI and NotI enzymes
and cloned into SalI/NotI site of pSPORT1 plasmid vector.
The DH10B E. coli host was transformed with the ligation
mixture by the standard chemical method. The average
insert size is about 3.0 kb. The library was constructed
by Yulan Piao (NIA)."
131 a 174 c 156 g 119 t

```

[illegible]

764 rGlnAspLysGlnLeuArgTyrSerAlaGlyValcIyAlaThrTrp...T T 780  
:::||||| |||||:::  
279 .....ATCCCGCTGGTCCTATGAGCAGCGCTGCCTCCGCCAC 243  
  
780 yrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeuAsnLys 796  
|||||::: |||||::: |||||:::  
242 TTGGCAGCATCGCTCGGTGGAGCTGAACACTACTGCATCTCCATTGGAGCGGTG 193  
  
797 LysGlnAsnAspGln...ThrAspThrValGlnPheGlnIleGlySerVa 812  
:::|||||::: ||| ||||| ||||  
192 CAGGGGGCGCACAGGATTGTGATGGTGCTCCAGTTTGGAGCTGGGATTTCG 143  
  
812 lPhe 813  
|||  
142 GTTC 139

seq\_name: gb\_estl:BB201163

seq\_documentation\_block:

LOCUS BB201163 641 bp mRNA linear EST 19-OCT-2001  
DEFINITION BB201163 RIKEN full-length enriched, 0 day neonate thymus Mus musculus CDNA clone A30034M11 3', mRNA sequence.

ACCESSION BB201163  
VERSION BB201163.2 GI:16271538  
KEYWORDS EST.  
SOURCE house mouse  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 641)  
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,  
Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,  
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,  
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,  
D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,  
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toyota,T.,  
Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)  
Unpublished (2001)

TITLE On Jun 30, 2000 this sequence version replaced gi:8866116.  
JOURNAL Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel.: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-resgsc.riken.go.jp,  
URL:http://genome.gsc.riken.go.jp/  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh  
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura  
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kiru,A. and  
Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara  
Y., and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaoka,I., Aizawa  
K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and  
Hayashizaki,Y.  
Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for  
further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

FEATURES  
source  
Location/Qualifiers  
1..641  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="A430034M11"  
/clone\_lib="RIKEN full-length enriched, 0 day neonate thymus"  
/tissue\_type="thymus"  
/dev\_stage="0 day neonate"  
/lab\_host="DH10B"  
/note="Site:1: Sali; Site:2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGATCCAGAGCTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCGAGTTAATAATATCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 129 a 170 c 190 g 152 t  
ORIGIN  
alignment\_scores:  
Quality: 131.00 Length: 199  
Ratio: 1.272 Gaps: 8  
Percent Similarity: 51.759 Percent Identity: 26.131  
alignment\_block:  
US-09-701-711-2 x BB201163 ..

Align seg 1/1 to: BB201163 from: 1 to: 641  
627 ArgProPheIleLysSerValLeuArgGlyTyrAlaLysLeuGlyTy 643  
15 AAGCGTGTGGCTTGGACTCGGTATTTCC.....ACGTCTCTCTGGGG 58  
643 rGlyAsnAsnLeuPro.....PheTyrGluAsnP 653  
59 TGAATGCTGTGGCCATCGGTGACAGCCATCCAGCATTTGTGACAGT 108  
653 heTyrAlaGlyTyrGlySerValArgGlyTyrAspLysSerLeu 669  
109 TTACTCTGGAGGCCCGACGAGTGTCCGAGGATTTAGCATGCACGCA 158  
670 GlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnThrTh 686  
159 GGACCCAGAGTGAA..... 173  
686 rLeuGlyGluValValGlyValGlyAsnAlaLeuAlaThrPheGlySerGluL 703  
174 ....GGAGATTACCTGGGGGGGAGCCCTACTGGGCTGGGGCCCTGCACC 219  
703 euIleLeuProLeuProPheLys.....GlyAspTyrIleAspGln 716  
220 TTACACCCCACTGCCCTCCGAGCCAGGCGGTGCTTCGAGAGCTT 269  
717 ValArgProValIlePheIleGlyGlyGlnValPheAspThrThrGl 733  
270 TTCAGACTCACTTTTCTCAATCGGGCAACCTG..... 305

733 yMetAspLysGlnThrIleAspLeuThrGlnPheLysAspProGlnAlaT 750  
306 .....TGCACCTCAACTATGTTGGTGGGGCCCCAAGCCC 339  
750 hrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuThrGlnAsp 766  
340 ATATCCGGAAGCTAGCTGAGTGC..... 362  
767 LysGlnLeuArgTyrSerAlaGlyValGlyAlaThrTrp...TyrThrPr 782  
363 .....ATCCGCTGTGCTATGGACGCGTCTCTCCGACTTGGCAA 406  
782 ofIleGlyProLeuSerIleSerTyrAlaLysProLeuAsnLysLysGlnA 799  
407 CATCGCTGGTGGAGCTGAAGTCTGCTATGGCGTGCAGGGGG 456  
799 snAspGln...ThrAspThrValGlnPheGlnIleGlySerValPhe 813  
457 GCGACAGGATTTGTGATGTGTCCAGTTTGGAGCTGGGATTCGGTTC 503

seq\_name: gb\_est2:BI539548

seq\_documentation\_block:  
LOCUS BI539548 548 bp mRNA linear EST 30-AUG-2001  
DEFINITION 452266 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.  
ACCESSION BI539548  
VERSION BI539548.1 GI:15380658

KEYWORDS EST.  
SOURCE cow.  
ORGANISM Bos taurus

REFERENCE  
AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T., Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett ,G.L., Heaton,M.P., Laegreid,W., Rohrer,G.A., Chitko-McKown,C.G., Perlea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and Keele,J.W.  
TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle

JOURNAL Genome Res. 11 (4), 626-630 (2001)

MEDLINE 21180013

COMMENT

Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called and alt trimmed with phred v0.980904.e. Vector identified by cross\_match with the -minscore 18 and -mismatch 12 options.  
PCR Primers  
FORWARD: AGGAACAGCTATGACCAT  
BACKWARD: GTTTCCTCCAGTCACGACG  
Plate: 114 row: B column: 1  
Seq primer: ATTTAGTGCACATATAG.  
Location/Qualifiers  
1..548  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="MARC 1BOV"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: pCMV SPORT6; Site:1: XbaI; Site:2: XhoI; Library made from pooled tissue from lymph node, ovary, fat, hypothalamus, and pituitary."

BASE COUNT 103 a 181 c 168 g 96 t  
ORIGIN

alignment\_scores:

Harvard University, Howard Hughes Medical Institute  
Department of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
MA 02138  
Tel: 617-495-1812  
Fax: 617-495-8557  
Email: dmelton@biohp.harvard.edu  
Pancreas was obtained from Gerard Gradwohl (PNAS 97 P1607-1611,  
2000) library was constructed by Catherine Lee DNA sequencing by:  
Washington University Genome Sequencing Center For information on  
obtaining a clone please contact: Dr. Marie Searce  
(mscearce@mail.med.upenn.edu)  
Seq primer: -400P from Gibco  
High quality sequence stop: 423.

FEATURES	source
Location/Qualifiers	
1..570	
/organism="Mus musculus"	
/strain="129/SV x CD1"	
/db_xref="taxon:10090"	
/clone_lib="Kaestner ngn3 wt"	
/dev_stage="p.c. 14.5"	
/db_host="E. coli-DH12S (GIBCO)"	
/note="Organ: pancreas; Vector: pSPORT1 (GIBCO); Site_1: Not I; Site_2: Sal I; The library was prepared by Catherine S. Lee and has not been published. The pancreas was obtained from Gerard Gradwohl (PNAS 97 P1607-1611, 2000). The cDNA's were prepared with an oligo containing a NotI site, and SalI linkers were added to the ends. The inserts were cut with NotI before being cloned into the NotI-SalI sites in the vectors. This is one of two libraries, ngn3 wt and ngn3 -/-. The wt library is in pSPORT1, T7 promoter is 5'."	
127 a	165 c 149 g 129 t
BASE COUNT	
ORIGIN	

```

alignment_scores:
  Quality: 130.50      Length: 168
  Ratio: 1.450         Gaps: 6
  Percent Similarity: 53.571  Percent Identity: 26.786

alignment_block:
  US-09-701-711-2 x BI714058/rev  ..

  Align seg 1/1 to reverse of: BI714058 from: 1 to: 570

651  GluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
555  GACAGGTTTACTTGGGAGGCCCCACGAGTGTCCGAGGATTAGCATGCA 506
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
667  rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGInG 684
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
505  CAGCATTGGACCCACAGCTGAA..... 484

684  lInThrLeuGlyGluValValGlyGlyAsnAlaLeuAlaThrPheGly 700
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
483  .....GGAGATTACTGGCGCGAGGCCCTACTGGGCTGGGGGC 445

701  SerGluLeuIleLeuProLeuProPheLys.....GlyAspTyrIle 714
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
444  CTGCACCTCTACACCCCTACTGCCCTTCGGCCAGGCCAGGGGGCTCGG 395

714  eAspGlnValArgProValIlePheIleGluGlyGlnValPheAspT 731
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
394  AGAGCTTTTCAGAACTCACTTTTCTCAATCGGGCAACCTG..... 352

731  hrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPheLysaspPro 747
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
351  .....TGCAACCTCAACTATGTCGAGGGCCCC 325

748  GlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuTh 764
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
324  AAAGCCCATATCCGGAAGCTAGCTCAGTGC..... 295

```

```

Quality: 130.50          Length: 166
Ratio: 1.483            Gaps: 6
Percent Similarity: 53.012 Percent Identity: 27.108

alignment_block:
US-09-701-711-2 x B1539548  ..

Align seg 1/1 to: B1539548 from: 1 to: 548

653 PheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSerSerLe 669
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2 TTCTACCTGGCGGACCCACCACCGTGGGGGCTTACGATGCACAGCGT 51

669 uGlyProArgSerGlnAlaTyrLeuThrAlaArgGlyGlnGlnThrT 686
:|||||:|||||:
52 CGGGCCACAGCGAA..... 67

686 hrLeuGlyGluValValGlyGlyAsnAlaLeuAlaThrPheGlySerGlu 702
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
68 .....GGTACATACCTGGCGGGGAAGCCTACTGGCGCGTGGCGTGCAC 112

703 LeuIleLeuProLeuProPheLys.....GlyAspTrpIleAspG1 716
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
113 CTGTACACCCGCTTACCTTCCGGCCGGCCAGGGCGCTTTGGCGAGCT 162

716 nValArgProValIlePheIleGluGlyGlyGlnValPheAspThrThrG 733
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
163 CTTTCAGGACACACTTCTTCTCAACGCGGGGAACCTC..... 199

733 LyMetAspLysGlnThrIleAspLeuThrGlnPheLysAspProGlnAla 749
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
200 .....TGCACCTCAACTACGGGGAGGGCCCCAAGGCC 232

750 ThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuThrGlnAs 766
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
233 CACATCCGGAAGCTGGCCGAGTGC..... 256

766 pLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrTrp...TyrThrP 782
:|||||:|||||: ||||| ||||| ||||| ||||| ||||| |||||
257 .....ATPCCGCTGTGCTTACGGGGCGGCATCGTCTCAGGCTGGGCA 299

782 roIleGlyProLeuSerIleSerTyrAlaLysProLeuAsnLysGln 798
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
300 ACATCGCGCGGCTGGAGCTCAACTACTGCGTCCCCATGGGGCTGACGCT 349

799 AsnAspGln...ThrAspThrValGlnPheGlnIleGlySerValPhe 813
:|||||: ||||| ||||| ||||| ||||| ||||| ||||| |||||
350 GGGGACAGATCTGTGACGCGGTCCTAGTTCGAGCTGGGATCAGGTTCC 397

```

seq_documentation_block:	
LOCUS	Bt714058 l3efilf12.xl Kaestner ngn3 wt Mus musculus cdna 3' similar to TR:09Y512 Q9Y512 DJ796117.2 ; mRNA sequence.
ACCESSION	Bt714058
VERSION	Bt714058.1 GI:15689753
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
DEFINITION	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 570)
AUTHORS	Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K., Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S., Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blinstead,A., Schmitt'A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams.T., Jackson,Y. and Bowers.Y.
TITLE	Endocrine Pancreas Consortium
JOURNAL	Unpublished (2000)
COMMENT	Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  Endocrine Pancreas Consortium



**TITLE** Systematic Analyses of NIA Mouse 7.5-dpc Whole Embryo cDNA Library (Long)

**JOURNAL** Unpublished (2001)

**COMMENT** Contact: Dawood B. Dudekula  
Laboratory of Genetics  
National Institute on Aging/National Institutes of Health  
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
Email: cdna@nigun.grc.nia.nih.gov  
Plate: C0283 row: B column: 03  
Seq primer: -21M13 Forward  
High quality sequence stop: 585  
POLYA-Yes.

**FEATURES**

source

Location/Qualifiers

1. 585

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taeef:C0283B03-3"

/db\_xref="taxon:10090"

/clone="C0283B03"

/clone\_lib="NIA Mouse 7.5-dpc Whole Embryo cDNA library (Long)"

/tissue\_type="whole embryo including extraembryonic tissues at 7.5-days postcoitum"

/dev\_stage="7.5-days postcoitum"

/lab\_host="DH10B"

/note="Vector: pSPORT1 (Invitrogen); Site\_1: SalI; Site\_2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://nigun.grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544191]). Total RNAs were extracted from a pool of four embryos at 7.5-days postcoitum. Double-stranded cDNAs were synthesized with an oligo(dT) primer [Invitrogen: 5'-pGACTAGTTCATCGCCGAGCGCCGCTTTTGTTTT-3'] from 7 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker LR-Sal4, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.2 kb. The library was constructed by Julian Piao 156 g 120 t

**BASE COUNT** 132 a 177 c 156 g 120 t

**ORIGIN**

alignment\_scores:

Quality: 130.50 Length: 168

Ratio: 1.450 Gaps: 6

Percent Similarity: 53.571 Percent Identity: 26.786

**alignment\_block:**

US-09-701-711-2 x BM205401/rev ..

Align seg 1/1 to reverse of: BM205401 from: 1 to: 585

651 GUAAGGTTTAACTGGGAGGCGCCACGAGTCTCCGAGGATTGACATCA 491

540 GACAGGTTTAACTGGGAGGCGCCACGAGTCTCCGAGGATTGACATCA 491

667 rSerLeuGlyProArSergAlaTyrLeuThrAlaArgGlyGing 684

490 CAGCATTTGACCCAGAGTGAA..... 469

684 InrThrLeuGlyAlaValAlaGlyAlaAlaAlaAlaAlaAlaAlaAlaAla 700

468 .....GGAGATTACTCTGGCGCGGCGCTACTGCGTGGGGG 430

701 SerGluLeuIleLeuProLeuProPheLys.....GlyAspTrpI 714

429 CTGACCTTACACCCACTGCTTCGGCCAGGCCAGGCGGCTGGC 380

714 eaSpGluValArgProValIlePheIleGluGlyGlyGluValPheAspT 731

379 AAGAGCTTTACAGACCACTTTTCTCTCAATGCGGCAACCTG..... 337

731 hrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPheLysAspPro 747

336 .....TGCAACCTCAACTGATGTGAGGCGCC 310

748 GlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuTh 764

309 AAGAGCCATATCCGAGAGCTAGCTAGTGC..... 280

764 rGlnAspLysGlnLeuArgTyrSerAlaGlyAlaAlaAlaAlaAlaAla 780

279 .....ATCCGCTGCTCTATGAGAGAGCGCTGCTCCTCGAC 243

780 YrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeuAsnLys 796

242 TTGGCACTCGCTCGCTGAGCTGAGCACTGCTGCTGCTGCTGCTGCTG 193

797 LysGlnAsnAspGln...ThrAspThrValGlnPheGlnIleGlySerVa 812

192 CAGGGGGGCGACAGAGATTGTGATGTGTCTTCAGCTTTGGAGCTGGATT 143

812 LpHe 813

142 GTTC 139

seq\_name: gb\_est1:A1553725

seq\_documentation\_block:

LOCUS A1553725 591 bp mRNA linear EST 12-MAY-1999

DEFINITION t28a04.x1 NCI-CGAP Brn25 Homo sapiens cDNA clone IMAGE:2168910 3'

similar to SM:YVP6.SCHPO Q10478 HYPOPHYTHICAL 51.8 KD PROTEIN

C17C9.06 IN CHROMOSOME 1.; mRNA sequence.

ACCESSION A1553725.1 GI:4486088

VERSION A1553725

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 591)

NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BRGAP), Tumor Gene Index

Unpublished (1998)

Contact: Robert Strausberg, Ph.D.

Email: cgaps-femail.nih.gov

Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www.bio.lnl.gov/bdip/image/image.html

Insert Length: 670 Std Error: 0.00

Seq primer: -40UP from G160

High quality sequence stop: 371

POLYA-No.

**FEATURES**

source

Location/Qualifiers

1. 591

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2168910"













```

/clone="AG-ND-17815"
/clone_lib="ND-TAM"
/notes="Vector: pECBAC1; Site_1: HindIII"
BASE COUNT      214 a      150 c      119 g      305 t
ORIGIN

alignment_scores:
  Quality: 130.50      Length: 267
  Ratio: 0.826      Gaps: 11
  Percent Similarity: 59.176      Percent Identity: 21.723

alignment_block:
US-09-701-711-2 x BH391826/rev ..

Align seg 1/1 to reverse of: BH391826 from: 1 to: 788

38  ILeThrGlyLeuGlnArgValThrIleGlnSerLeuGlnSerValLeuPr 54
   :::::::::::::::::::: ||| |||::::::::::
785 GTGATGCTTAAAAAATACACACCTGACCAATTTTACGCTTACAGG 736
   :::::::::::::::::::: ||| |||::::::::::
54  oPheArgLeuGlyGlnValSer.....GluAsnGlnLeuAlaAspG 69
   :::::::::::::::::::: ||| |||::::::::::
735 TCTTGTAAGAGTGAATCTGTGAATCCCGGACAAAGACTGAGTACAG 686
   :::::::::::::::::::: ||| |||::::::::::
69  lyVallyAlaLeuTyraIleThrGlyAsnPheserAspValGlnValTy 85
   :::::::::::::::::::: ||| |||::::::::::
685 CTATCAAAAAGCTTGGGATAGCCAAATATTTTCTGAAGTTAAGTATAT 636
   :::::::::::::::::::: ||| |||::::::::::
86  HisGln.....GluGlyArgIleIle.....TyrGlnValThrG 97
   :::::::::::::::::::: ||| |||::::::::::
635 GTGCAGAGCATGTGAGGACAAATATCTGTTAAATCTCTCTTACGA 586
   :::::::::::::::::::: ||| |||::::::::::
97  uArgProLeuIleAlaGluIleAsnPhesGluGlyAsnArgLeuIlePro 114
   :::::::::::::::::::: ||| |||::::::::::
585 TTAAAAAGAACTGTGTAAGTAAATTTTACCGGA...AAGGAATCAAAA 539
   :::::::::::::::::::: ||| |||::::::::::
114 yAsGluGlyLeuGlnGluGlyLeuLysAsnLagIleuAlaVal.GlyG 130
   :::::::::::::::::::: ||| |||::::::::::
538 AATCTAAAAAGCAAACTTATTAAGATTAATACCTGAAAGCTGTGAT 489
   :::::::::::::::::::: ||| |||::::::::::
130 nProLeuLysGlnAlaThrValGlnMetIleGluThrGluLeuThrAsn. 146
   :::::::::::::::::::: ||| |||::::::::::
488 GAAATTTACCGACCGAGAACTGTTACTAATCACTAATGATGACCTCAG 439
   :::::::::::::::::::: ||| |||::::::::::
147 GlnTyrIleSerGlnGlyTyrTyrAsnThrGluIleThrVal..... 161
   :::::::::::::::::::: ||| |||::::::::::
438 CAATACATTAACAAGAGATCCCGATGCTAAGATCACTTGAAGATAA 389
   :::::::::::::::::::: ||| |||::::::::::
161 sGlnThrMetLeuAspGlyAsnArgValLysLeuAspMetThrPheAlaG 178
   :::::::::::::::::::: ||| |||::::::::::
388 AATCAATGCAAGAGATCTAGCTTATCGACTGACCATTAATGATAGATA 339
   :::::::::::::::::::: ||| |||::::::::::
178 lGluLysProAlaArgValAlaAspIleAsnIleIleGlyAsnGlnHis 194
   :::::::::::::::::::: ||| |||::::::::::
338 AAGGAAACCTGTTAAGATGACCGATGATTTTGAAGAAACACAGT 289
   :::::::::::::::::::: ||| |||::::::::::
195 PheserAspAlaAspLeuIleAspValLeuAlaIleLysAsnLys.. 210
   :::::::::::::::::::: ||| |||::::::::::
288 GTTTCCTCACTAACT...CGTAAACGCTTTTAAAAATCGAACA 242
   :::::::::::::::::::: ||| |||::::::::::
211 .....IleAsnProLeuSerLysAlaAspArgTyrThrGlnGlu 224
   :::::::::::::::::::: ||| |||::::::::::
241 AAGAGATTCCTCTGTTGCTTAAAGCTTGAAATTTATCAAGATA 192
   :::::::::::::::::::: ||| |||::::::::::
224 ySLeuValThrSerLeuGluAsnLeuArgAlaLysTyrLeuAsnAlaGly 240
   :::::::::::::::::::: ||| |||::::::::::
191 AATATGAGAGAGATGAGACCTCTGTAGACTACACTCAACTCTTTGGA 142
   :::::::::::::::::::: ||| |||::::::::::
241 PheValArgPheGluIleLysAspAlaLysLeu.....AsnIleAs 254
   :::::::::::::::::::: ||| |||::::::::::
141 TTT.....AGAGATATGAGAGTAGTGTCAGACTCTGTTTC 107

```

```

254  nGluAspLysAsnArgIlePheValGlnIleSerLeuHisGluGlyGluG 271
   :::::::::::::::::::: ||| |||::::::::::
106 GCGTAAATCTAAGCATACACATCAAAAGTAAATAGATAGAGGTAATA 57
   :::::::::::::::::::: ||| |||::::::::::
271  lNTrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuThrTyrThr 286
   :::::::::::::::::::: ||| |||::::::::::
56  AATACTATATCGAGATATTTACTCTGTAGGAAATACAGTATTTTCT 10
   :::::::::::::::::::: ||| |||::::::::::

seq_name: gb.est1:AM107426

seq_documentation_block:
LOCUS      AM107426                871 bp      mRNA      linear      EST 20-Oct-1999
DEFINITION un16c03.x1 Sugano mouse kidney mKia Mus musculus cDNA clone
IMAGE:2192452.3' similar to WP:034E10.1 CE01181 ;, mRNA sequence.
ACCESSION  AM107426
VERSION    AM107426.1  GI:6078226
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 871)
AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,R., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Page,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
            The WashU-NCI Mouse EST Project 1999
            Unpublished (1999)
            Other_ESTs: un16c03.y1
            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@wustl.wustl.edu
            This clone is available royalty-free through LNLN; contact the
            IMAG Consortium (info@image.llnl.gov) for further information.
            MGI:1004904
            Seq primer: custom primer used
            High quality sequence stop: 461.
            Location/Qualifiers
                1..871
                /organism="Mus musculus"
                /strain="C57BL"
                /db_xref="taxon:10090"
                /clone="IMAGE:2192452"
                /clone_lib="Sugano mouse kidney mKia"
                /sex="female"
                /dev_stage="adult"
                /lab_host="DH10B"
                /note="Organ: Kidney; Vector: pME18S-FL3; Site_1: DraIII
                (CACTGTGG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                was primed with an oligo(dT) primer
                (ATGGGCGCTTTTCTTTTCTTTT); double-stranded cDNA was
                ligated to a DraIII adaptor [TGTGGCGCTTACGG], digested
                and cloned into distinct DraIII sites of the pME18S-FL3
                vector (3' site CACTGTGG, 3' site CACCATGTG). XhoI should
                be used to isolate the cDNA insert. Size selection was
                performed to exclude fragments <1.5kb. Library
                constructed by Dr. Sumio Sugano (University of Tokyo
                Institute of Medical Science). Custom primers for
                sequencing: 5' end primer CTCTGCTCTTAAAGCTGGC and 3' end
                primer CGACTGCGCTGAGCACA."
BASE COUNT      201 a      254 c      227 g      181 t      8 others
ORIGIN

alignment_scores:
  Quality: 130.50      Length: 168
  Ratio: 1.450      Gaps: 6
  Percent Similarity: 53.571      Percent Identity: 26.786

```

## alignment\_block:

US-09-701-711-2 x AM107426/rev ..

Align seg 1/1 to reverse of: AM107426 from: 1 to: 871

```

651  GLUAsnpheTYrAlaGlyTYrGlySerValaArgGlyTYrAspGlnse 667
      ::::::::::::::::::::|
538  GACAGGTTTACCTGGAGGCCACGAGGTGTCCGAGGATTACCATCA 489
      ::::::::::::::::::::|
667  rSerleuGlyProArGSerGlnAlaTYrLeuThraAlaArgArgGlyGln 684
      ::::::::::::::::::::|
488  CAGCATGTGACCCGACAGTGAA..... 467
      ::::::::::::::::::::|
684  InThrThrLeuGlyGluValaGlyGlnAlaLeuAlaThrPheGly 700
      ::::::::::::::::::::|
466  .....GGAGATTACCTGGCGCGCGGCTTACTGGCGTGGGGGC 428
      ::::::::::::::::::::|
701  SerGluLeuLeuProLeuProPheLys.....GlyAspTPII 714
      ::::::::::::::::::::|
427  CTGCACCTCTACACCCACCTGCTCCGGCCAGCCAGGCTGCTTCGG 378
      ::::::::::::::::::::|
714  eAspGluValaArgProValaIlePheIleGluGlyGlnValaPheAsp 731
      ::::::::::::::::::::|
377  AGACCTTTTACGAACTCTTTTCTCAATGCGGCAACCTG..... 335
      ::::::::::::::::::::|
731  hrThrglyMetAspLysGlnThrIleAspLeuThrglnPheLysAspPro 747
      ::::::::::::::::::::|
334  .....TGCNACCTCAACTATGTTGGTGGGGGCC 308
      ::::::::::::::::::::|
748  GlnAlaThraIaGluGlnAsnAlaLysAlaAlaAsnArgProLeuThr 764
      ::::::::::::::::::::|
307  AAGCCCATATCCGAGCAAGCTAGCTGATC..... 278
      ::::::::::::::::::::|
764  rGlnAspLysGlnLeuArgTYrSerAlaGlyAlaGlyAlaThrTP...T 780
      ::::::::::::::::::::|
277  .....ATCCGCTGTCTTATGAGCAGCGCTGCTCCGAC 241
      ::::::::::::::::::::|
780  yrThrProIleGlyProLeuSerIleSerTYrAlaLysProLeuAsnLys 796
      ::::::::::::::::::::|
240  TTGGCAACATCCGCTGCTGAGCTGACACTGCACTTCCTATGGCGGTG 191
      ::::::::::::::::::::|
797  LysGlnAsnAspGln...ThraSPThrValGlnPheGlnIleGlySerVa 812
      ::::::::::::::::::::|
190  CAGGGGGGGCAGACGATTTGTGTGTTGTCAGATTGGAGCTGGAGTTCG 141
      ::::::::::::::::::::|
812  lPhe 813
      ::::::::::::::::::::|
140  GTTC 137

```

seq\_name: gb\_est2:BM233588

seq\_documentation\_block:

LOCUS BM233588 580 bp mRNA linear EST 31-JAN-2002

DEFINITION K0338H12-3 NIA Mouse Osteoblast cDNA Library (Long) Mus musculus

ACCESSION BM233588

VERSION BM233588.1 GI:17868858

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryote: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.

REFERENCE 1 (bases 1 to 580)

AUTHORS Plao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Luo, A., Carter, M.G., Umezawa, A. and Ko, M.S.H.

TITLE Systematic Analyses of NIA Mouse Osteoblast cDNA Library (Long)

JOURNAL Unpublished (2001)

COMMENT Contact: Dawood B. Dudekula

Laboratory of Genetics

National Institute on Aging/National Institutes of Health

333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA

Email: cdna@igsun.grc.nia.nih.gov

Plate: K0338 row: H column: 12

## FEATURES

Seq primer: -21M13 Forward

High quality sequence stop: 580

POLY-A-Tes.

## source

Location/Qualifiers

1..580

/organism="Mus musculus"

/strain="C3H/He mice"

/db\_xref="niaEST:K0338H12-3"

/db\_xref="taxon:10090"

/clone="K0338H12"

/clone\_1ib="NIA Mouse Osteoblast cDNA Library (Long)"

/tissue\_type="Osteoblast"

/cell\_line="KUSA-A1 cells"

/lab\_host="DH10B"

/note="vector: pSPORT1 (Invitrogen); Site\_1: SalI; Site\_2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were obtained from Dr. Akihito Umezawa (Keio University School of Medicine, Japan). Double-stranded cDNAs were synthesized with an Oligo(dt) primer [Invitrogen: 5'-DCACTAGTTCTAGATCGCAGCGCGCCCTTTTCTTTT-3'] from 2.1 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker IL-SalI, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Tag polymerase (Takara) with a primer SalI-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 3.0 Kb. The library was constructed by Yulan Plao (NIA)."

## BASE COUNT

131 a 174 c 157 g 118 t

## alignment\_scores:

Quality: 129.50 Length: 168

Ratio: 1.439 Gaps: 6

Percent Similarity: 53.571 Percent Identity: 26.786

## alignment\_block:

US-09-701-711-2 x BM233588/rev ..

Align seg 1/1 to reverse of: BM233588 from: 1 to: 580

```

651  GLUAsnpheTYrAlaGlyTYrGlySerValaArgGlyTYrAspGlnse 667
      ::::::::::::::::::::|
540  GACAGGTTTACCTGGAGGCCACGAGGTGTCCGAGGATTACCATCA 491
      ::::::::::::::::::::|
667  rSerleuGlyProArGSerGlnAlaTYrLeuThraAlaArgArgGlyGln 684
      ::::::::::::::::::::|
490  CAGCATGTGACCCGACAGTGAA..... 469
      ::::::::::::::::::::|
684  InThrThrLeuGlyGluValaGlyGlnAlaLeuAlaThrPheGly 700
      ::::::::::::::::::::|
468  .....GGAGATTACCTGGCGCGCGGCTTACTGGCGTGGGGGC 430
      ::::::::::::::::::::|
701  SerGluLeuLeuProLeuProPheLys.....GlyAspTPII 714
      ::::::::::::::::::::|
429  CTGCACCTCTACACCCACCTGCTCCGGCCAGCCAGGCTGCTTCGG 380
      ::::::::::::::::::::|
714  eAspGluValaArgProValaIlePheIleGluGlyGlnValaPheAsp 731
      ::::::::::::::::::::|
379  AGACCTTTTACGAACTCTTTTCTCAATGCGGCAACCTG..... 337
      ::::::::::::::::::::|
731  hrThrglyMetAspLysGlnThrIleAspLeuThrglnPheLysAspPro 747
      ::::::::::::::::::::|

```



SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
AUTHORS 1 (bases 1 to 541)  
TITLE Bonaldo,M.F., Lennon,G. and Soares,M.B.  
JOURNAL Normalization and subtraction: two approaches to facilitate gene discovery  
MEDLINE Genome Res. 6 (9), 791-806 (1996)  
COMMENT 97044477  
Contact: Chin, H  
National Institute of Mental Health  
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
20892-9643, USA  
Tel: 301 443 1706  
Fax: 301 443 9890  
Email: mestr@mail.nih.gov  
CDNA Library Preparation: M.B. Soares Lab Clone distribution:  
Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements  
Seq primer: M13 Reverse

FEATURES  
Source  
Location/Qualifiers  
1..541  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UI-M-BH1-ame-c-09-0-UT"  
/clone\_1db="NIH-BMAP\_M\_S2"  
/dev\_stage="27-32 days"  
/lab\_host="DH10B (Life Technologies)"  
/note="vector: pUT3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; The NIH-BMAP\_M\_S2 library is a subtracted library derived from NIH-BMAP\_M\_S1, which in turn is a subtracted library derived from a mixture of normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hippocampus). The driver used for subtraction consisted of a pool of 5,000 clones from the NIH-BMAP\_M\_S1 library and a pool of 2,000 clones obtained from non-normalized and normalized mouse brain spinal cord libraries."  
BASE COUNT 106 a 148 c 162 g 125 t  
ORIGIN

alignment\_scores:  
Quality: 124.50 Length: 168  
Ratio: 1.383 Caps: 6  
Percent Similarity: 53.571 Percent Identity: 26.190

alignment\_block:  
US-09-701-711-2 x BE647176 ..

Align seg 1/1 to: BE647176 from: 1 to: 541

```

651 GUAANPHEtYtAlAGlYtYrGlYSeRvAlArGtYtYrASpGlSe 667
      ::::::::::::::::::::|
132 GACAGGTTTACCTGGAGGCCCGCCAGAGTGTCCGAGATTATGACATGCA 181
      |||:|||||:|||||:
667 rSerLeuGlYProArGSerGlnAlATyTLeuThrAlArGArGtGtGtG 684
      |||:|||||:|||||:
182 CAGCATTTGACCCCGAGAGTGA..... 203
      |||:|||||:|||||:
684 InTtHrThLeuGlYAluValAGlYgLYASnAlAlaLeuAlAThrPhGtY 700
      |||:|||||:|||||:
204 .....GGAGATTACTGGCGCGGAGGCTTACTGGGCTGGGCGC 242
      |||:|||||:|||||:
701 SerGluLeuLeuLeuProLeuProPhelys.....GlyASPTpTl 714
      |||:|||||:|||||:

```

```

243 CTGCACCTCTACACCCACCTCCCGGCCAGGCCAGGGTGCTTCG 292
714 eASpGlnValArProValAllePheIlleGlYgLYAluValPheASpT 731
      |||:|||||:|||||:
293 AGACCTTTTCAGACCTCATTTTCCATTCGCGGACACCTG..... 335
731 hTtHrGlYMetASpLYGlnThrIleASpLeuThrGlnPhelyASpPro 747
336 .....TGCACCTCAACTATGTGTGGAGGCCGCC 362
748 GlnAlAThrAlAGlYgLYASnAlAlaAlASnArGrProLeuLeuTh 764
      |||:|||||:|||||:
363 AAGGCCATTTCCGAGACCTAGCTGATGC..... 392
764 rGlnASpLYGlnLeuArGrTYSerAlAGlYAluValAThrTTP...T 780
      |||:|||||:|||||:
393 .....ATCCGCTGTCTCTATGGAGAACGGCTGCTCCGAC 429
780 YrTtHrProIlleGlYProLeuSerIlleSerYrAlASpProLeuASn 796
      |||:|||||:|||||:
430 TTGGCAACATTCGCTCGGCTGGACCTGACATTCGATTCCTATGGCG 479
797 LysGlnASnASpGln...ThrASpThrValGlnPhGlnIlleGlySer 812
      |||:|||||:|||||:
480 CAGGGGGGCGACAGATTGTGTATGTGTCCACTTTGGAGCTGGATTTC 529
812 lPhe 813
530 GtTC 533
seq_name: gb_est1:AM027570

```

seq documentation block:  
LOCUS AM027570 596 bp mRNA linear EST 09-MAR-2000  
DEFINITION w74h01.x1 Soares\_thymus.NHFT Homo sapiens cDNA clone  
IMAGE:2535313 3' similar to SW:XF6.SCH0 Q10478 HYPOTHETICAL 51.8  
KD PROTEIN C173.06 IN CHROMOSOME 1.; mRNA sequence.  
ACCESSION AM027570 GI:5886326  
VERSION AM027570.1 GI:5886326  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 596)  
NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaps@email.nih.gov  
NOTE This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert length: 1344 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 457.  
FEATURES  
Location/Qualifiers  
1..596  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2535313"  
/clone\_1db="Soares\_thymus.NHFT"  
/dev\_stage="fetal"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: thymus, pooled; Vector: pUT3D-Pac  
(Pharmacia) with a modified polylinker; Site\_1: Not I;  
Site\_2: Eco RI; 1st strand cDNA was primed with a Not I -  
oligo(dT) primer [5'  
TGTTACCAATCTGAGTGGAGCGCGCCGCAAGCTTTTCTTTTCTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pUT3 vector. Library  
went through one round of normalization. Library



constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 124 a 171 c 165 g 136 t

ORIGIN

alignment\_scores:

Quality: 124.50

Length: 168

Ratio: 1.415

Gaps: 6

Percent Similarity: 52.381 Percent Identity: 26.190

alignment\_block:

US-09-701-711-2 x AM027570/rev ..

Align seg 1/1 to reverse of: AM027570 from: 1 to: 596

```

651 GUASNPHEYYALAGLYTYRGLYSERVALARGLYTYRASPGLNSE 667
      ::::::::::::::::::::|
554 GATAGGTTTACCTCGGGGACCCACAGCGTCCGGATTCCAGCATGCA 505
      ::::::::::::::::::::|
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
      |||::|
504 CAGCATCGGGCCACAGACGAA..... 483
      |||::|
684 InThrThrLeuGlyGluValAlaGlyGlyAsnAlaLeuAlaThrPheGly 700
      |||::|
482 .....GGAGACTACTAGTGAAGAGAGCTACTGGGCGCGCGGC 444
      ::::::::::::::::::::|
701 SerGluLeuIleuProLeuProPheLys.....GlyAspTyrP11 714
      ::::::::::::::::::::|
443 CTGCACCCCTTACCCCTTACCTTCCGGCCAGGCCAGCGGCGCTTGG 394
      ::::::::::::::::::::|
714 eAspGlnValArgProValIlePheIleGlyGlyGlnValPheAspT 731
      ::|
393 AGAAGCTTTCCGACACACACTTCTTCTCAGCGAGAAACCTC..... 351
      ::|
731 InThrGlyMetAspLysGlnThrIleAspLeuThrGlnPheLysAspPro 747
      ::|
350 .....TGCAACCTCACTATGCGGAGGAGGCGCC 324
      ::|
748 GlnAlaThrAlaGlnGlnAsnAlaLysAlaAlaAsnArgProLeuLeuTh 764
      ::|
323 AAAGCTCATATTCTGAAGCTGCTGAGTGC..... 294
      ::|
764 rGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrTrp...T 780
      ::|
293 .....ATCCGCTGCTGACGGGCGCGGATGTCCTCAGGC 257
      ::|
780 YrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeuAsnLys 796
      |||::|
256 TTGCAACATCGCTCGTTGGACTTATCTGCGTCCCATGGAGTA 207
      ::|
797 LysGlnAsnAspGln...ThrAspThrValGlnPheGlnIleGlySerVa 812
      ::|
206 CAGCAGAGGTGACAGGATATGTATGCTGCCATTGGAGCTGGGATAG 157
      ::|
812 Lphe 813
      |||
156 GTTC 153

```

seq\_name: gb\_est1:AM65916

seq\_documentation\_block:

LOCUS AM65916 598 bp mRNA linear EST 06-APR-2000

DEFINITION h195f07.x1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone

IMAGE:2980069.3' similar to SW:YDF6\_Q10478 HYPOTHEETICAL 51.8

CD PROTEIN C17C9.06 in CHROMOSOME 1. / mRNA sequence.

ACCESSION AM65916

VERSION AM65916.1 GI:7458465

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 598)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

JOURNAL Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaps-remail.nih.gov

COMMENT This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([infoimage.lnl.gov](http://infoimage.lnl.gov)) for further information.

Seq primer: -400P from Gluco  
High quality sequence stop: 429.

FEATURES

source

1..598

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2980069"

/clone\_lib="Soares\_NFL\_T\_GBC\_S1"

/lab\_host="DH10B"

/note="Organ: pooled; Vector: pT7/3D-Pac (Pharmacia) with

a modified polylinker; Site.1: Not I; Site.2: Eco RI;

Equal amounts of plasmid DNA from three normalized

libraries (fetal lung NDHL19W, testis NHT, and B-cell

NCI-CGAP\_GCB1) were mixed, and ss circles were made in

vitro. Following HAP purification, this DNA was used as

tracer in a subtractive hybridization reaction. The driver

was PCR-amplified cDNAs from pools of 5,000 clones made

from the same 3 libraries. The pools consisted of

I.M.A.G.E. clones 297480-302087, 682632-687239,

726408-728711, and 729096-731399. Subtraction by Bento

Soares and M. Fatima Bonaldo.

BASE COUNT 125 a 170 c 163 g 140 t

ORIGIN

alignment\_scores:

Quality: 123.50

Length: 168

Ratio: 1.403

Gaps: 6

Percent Similarity: 52.381

Percent Identity: 26.190

alignment\_block:

US-09-701-711-2 x AM65916/rev ..

Align seg 1/1 to reverse of: AM65916 from: 1 to: 598

```

651 GUASNPHEYYALAGLYTYRGLYSERVALARGLYTYRASPGLNSE 667
      ::::::::::::::::::::|
556 GATAGGTTTACTTCGGGTACCCACAGCGTCCGGGATTCAGCATGCA 507
      ::::::::::::::::::::|
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
      |||::|
506 CAGCATCGGGCCACAGACGAA..... 485
      |||::|
684 InThrThrLeuGlyGluValAlaGlyGlyAsnAlaLeuAlaThrPheGly 700
      |||::|
484 .....GGAGACTACTAGTGAAGAGAGCTACTGGGCGCGCGGC 446
      ::|
701 SerGluLeuIleuProLeuProPheLys.....GlyAspTyrP11 714
      ::|
445 CTGCACCTTACACCCCATTTCTCCGCGCAGGCCAGCGGCGCTTGG 396
      ::|
714 eAspGlnValArgProValIlePheIleGlyGlyGlnValPheAspT 731
      ::|
395 AGAAGCTTTCCGACACACACTTCTTCTCAGCGAGAAACCTC..... 353
      ::|
731 InThrGlyMetAspLysGlnThrIleAspLeuThrGlnPheLysAspPro 747
      ::|
352 .....TGCAACCTCACTATGCGGAGGAGGAGGCGCC 326
      ::|
748 GlnAlaThrAlaGlnGlnAsnAlaLysAlaAlaAsnArgProLeuLeuTh 764
      ::|
325 AAAGCTCATATTCTGAAGCTGCTGAGTGC..... 296
      ::|
764 rGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrTrp...T 780

```







Sequence	Strd Orig	ZScore	EScore	Len	! Documentation
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA2000.DAT:AAZ29550 +				813.00	14729.29
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA2000.DAT:AAZ29551 +				749.00	13567.10
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA2001A.DAT:AAK28550 +				749.00	13540.49
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA2001A.DAT:AAK28551 +				36.00	625.99
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA2000.DAT:AAZ29554 +				11.00	192.08
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA2000.DAT:AAZ29555 +				11.00	191.40
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA2001A.DAT:AAK60366 +				10.00	149.51
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA2001B.DAT:AAZ58185 +				10.00	144.37
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA1997.DAT:AAAT94180 -				9.00	135.69
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA1998.DAT:AAV53520 -				9.00	135.65
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA2001A.DAT:AAK61065 +				9.00	130.65
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA2001A.DAT:AAI229004 +				8.00	139.03
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA2001A.DAT:AAK60156 +				8.00	129.93
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA2001A.DAT:AAK60152 +				8.00	124.32
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA2000.DAT:AAZ56035 -				8.00	123.88
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA2000.DAT:AAAC03891 +				8.00	123.31
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA1994.DAT:AAO79437 -				8.00	123.26
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA2001A.DAT:AAK62517 -				8.00	122.35
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA2001A.DAT:ABA29847 -				8.00	122.35
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA2001A.DAT:AAK10872 -				8.00	122.35
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA2001A.DAT:AAK36734 -				8.00	122.35
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA2001A.DAT:AAI17584 -				8.00	122.35
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA2001A.DAT:AAI42500 -				8.00	122.35
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA2001A.DAT:AAZ42398 +				8.00	121.40
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA1998.DAT:AAV35370 +				8.00	121.15
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA1999.DAT:AAZ32098 +				8.00	121.15
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA2000.DAT:AAK16109 +				8.00	118.76
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA2001B.DAT:AAAS18139 +				8.00	118.59
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA1998.DAT:AAV35383 +				8.00	117.35
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA1999.DAT:AAZ32093 +				8.00	117.35
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA2000.DAT:AAK38511 -				8.00	116.57
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA1993.DAT:AAK38956 +				8.00	113.99
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA2001A.DAT:AAH74178 +				8.00	112.82
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA2000.DAT:AACT78114 -				8.00	112.81
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA2001B.DAT:ABLO5175 -				8.00	112.53
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA2001B.DAT:AAAS81841 +				8.00	112.36
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA2001B.DAT:ABU24790 -				8.00	111.43
/SIDS1/gcgdata/geneseg/geneseqn-emb1/NA2001A.DAT:AAAD07053 +				8.00	111.35

Align seg 1/1 to: AAZ29550 from: 1 to: 2442

1	MetArgAsnSerTyrPheLysGlyPheGlnValSerAlaMetThrMetAl	17
1	ATGGCTAATTATCATATTTAAAGGTTTTCAGGTTCAGTGC	50
17	aValMetMetValMetSerThrHisAlaGlnAlaAlaAspPheMetAlaA	34
51	TGTCATGATGGTAAGTCAACTCATGCACAAGCGCGGATTTATGGCAA	100
34	snAspIleThrIleThrGlyLeuGlnArgValThrIleGluSerLeuGln	50
101	ATGACATTACCATCACAGACTACAGCGAGTGACCATTCGAAAGCTTACA	150
51	SerValLeuProPheArgLeuGlyGlnValValSerGluAsnGlnLeuAl	67
151	AGCGTGTCCGCTTTTCGCTTGGTCAAGTGTGTGAGCGAAGAACAGTGGC	200
67	aAspGlyValLysAlaLeuTyrAlaThrGlyAsnPheSerAspValGlnV	84
201	TGATGGTGTCAAGCACCTTATGCACAGGCATTTTTCAGATGTGCAG	250
84	alTyrHisGlnGluGlyArgIleIleTyrGlnValThrGluArgProLeu	100
251	TCTATCATCAAGAGGGCGTATCATCTATCAGTATACCGCAAGGCGGT	300
101	IleAlaGluIleAsnPheGluGlyAsnArgLeuIleProLysGluGlyLe	117
301	ATCGCTGAGATTAAATTTTGAGGGCAATCGCTTAATTCGCAAGAGGTCT	350
117	uGlnGluGlyLeuLysAsnAlaGlyLeuAlaValGlyGlnProLeuLysG	134
351	ACAAGAGGGCTAAAAATGCTGGCTTAGCTGTGGGTCAACCATTAAC	400
134	lnAlaThrValGlnMetIleGluThrGluLeuThrAsnGlnTyrIleSer	150
401	AAGCCACAGTACAGATGATCGAAACCGAGCTTACCAATCAATATATCA	450
151	GlnGlyTyrTyrAsnThrGluIleThrValLysGlnThrMetLeuAspG	167
451	CAAGCTATTATATATCCGAAATTTACTGCTCAACAGACATGCTTGATGG	500
167	yAsnArgValLysLeuAspMetThrPheAlaGluGlyLysProAlaArgV	184
501	TAATCGTGTAAAGCTTGATATGACCTTTGCTGAAGGTAAACCTGCACGG	550
184	alValAspIleAsnIleGlyAsnGlnHisPheSerAspAlaAspLeu	200
551	TGGTTGATTTATATCATCGCAATCAGCATTTTAGCGATGCAGATTG	600
201	IleAspValLeuAlaIleLysAspAsnLysIleAsnProLeuSerLysAl	217
601	ATTGATGTGCTTGCAGTAAAGGATAATAAATCAATCCACTGCTCAAGC	650
217	aAspArgTyrThrGlnGluLysLeuValThrSerLeuGluAsnLeuArgA	234
651	TGACCGTTTACTCAAGAAAGCTGGTGACCATTTTAGAGAAATTTGCGTG	700
234	laLysTyrLeuAsnAlaGlyPheValArgPheGluIleLysAspAlaLys	250
701	CTAAATATCTCAATGCAGGCTTTGCGCTTTTGAGATTAAAGATGCTAAG	750
251	LeuAsnIleAsnGluAspLysAsnArgIlePheValGluIleSerLeuHi	267
751	CTTAATATTAATGAAGATAAAACCGTATCTTTGTGTGAGATTTCATTGCA	800
267	sGluGlyGluGlnTyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuT	284
801	TGAAGGTGACCAATATCGCTTTTGACACACACAGTTTTTGGTATATTAA	850
284	hrTyrThrGlnAlaGluLeuGluAlaLeuLeuLysPheLysAlaGluGlu	300
851	CTTATCTCAAGCAGAACTTGAGCACTGCTTAATTTCAAGCGCAAGAA	900

301	GlyPheSerGlnAlaMetLeuGlnThrThrAsnAsnIleSerThrLy	317
901	GGGTTTTTACAAAGCCATGCTTGAGCAAAACAACAACAATATCATGTACCA	950
317	sPheGlyAspaspGlyTyrTyrTyrAlaGlnIleArgProValThrArgI	334
951	ATTTCGGTGACGATGGCTATTATTATGTCTCAATCCGTCCTGTACACGCA	1000
334	IleAsnAspGluSerArgThrValAspValGluTyrTyrIleAspProVal	350
1001	TTAATGATGAAGTCGTACGGTTGATGTGGAAATATATATTGACCCCTGA	1050
351	HisProValTyrValArgArgIleAsnPheThrGlyAsnPheLysThrGl	367
1051	CACCCTGCTATGTACGCCGATTAATTTTACAGGTACTTTTAAGAACCCA	1100
367	nAspGluValLeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlas	384
1101	AGATCAGGTACTCCGTCGTGAGATGCGACAACCTTGAAAGTCGGTTGGCAT	1150
384	erAsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhe	400
1151	CTAATCAAAAAATCCAGCTGCTCGTGACGCTTGATGCGGACTGGGT	1200
401	PheLysHisValThrValAspThrArgProValProAsnSerProAspGl	417
1201	TTTAAACATGTTACCGTTGATGTACGTGCTCCAGTACCCCACTCACCTGATCA	1250
417	nValAspValAsnPheValValGluGlnProSerGlySerSerThrI	434
1251	GCTTCATGTAAATTTTGGGTTGAAGACAACCTTCAGGATCATCAACCA	1300
434	IeAlaIleGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspVal	450
1301	TCGCAGCAGGCTACTCTCAAAAGTGGTGGTGAACCTTTTCAATTTGATGTT	1350
451	SerGlnAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSe	467
1351	TCTCAAAATTAACCTTATGGGTACAGGTAAAGCACGTCAATGCTTCGTGTTTC	1400
467	rArgSerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrP	484
1401	TCGCTCTCAGACCCGCTGAGGTGTATAGTTTGGGTATGACCAACCCATACT	1450
484	heThrValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThr	500
1451	TTACCGTAAATGGGCTCTCGCAAAAGCTTGAGTGCTACTATPCGCAAAACC	1500
501	LysTyrAspAsnLysAsnIleSerAsnTyrValLeuAspSerTyrGlyGl	517
1501	AAGTATGATAACAAGAACATTAGTAATATGTACTTCATTGTTGCTGGTGG	1550
517	ySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSerPheG	534
1551	CTCATTAAGCTATGGATATCCAATTTGATGAATAAATCAACGATGAAGCTTG	1600
534	lyLeuAsnAlaAspAsnThrLysLeuHisGlyGlyArgPheMetGlyIle	550
1601	GCTCGAATGCTGACAAATACCAAGCTTCATGGCGGTCTGTTTTATGGGCATT	1650
551	SerAsnValLysGlnLeuMetAlaAspGlyGlyLysIleGlnValAspAs	567
1651	AGTAATGTCAACGACGCTCATGGCAGATGGTGGCAAAATTCAAAGTGGATAA	1700
567	nAsnGlyIleProAspPheLysHisAspTyrThrTyrAsnAlaIleI	584
1701	TAATGGCATTCTCGATTTTAAAGCAIGATTACACAACCTACATGCCATT	1750
584	euGlyTyrAsnTyrSerSerLeuAspArgProValPheProThrGlnGly	600
1751	TGGGGTGGAAATATTCAGTCTAGATCGCTCGCTGTATTTTCCAAACCCAAAGC	1800

601 MetSerHisSerValAspLeuThrValGlyPheGlyAspLysThrHisG1 617  
 |||||  
 1801 ATGAGTCATCTGTAGATTGACGGTTGGTTTGGTGATAAACTCATCA 1850  
 |||||  
 617 nLysValValTyrGlnGlyAsnIleTyrArgProPheIleLysLysServ 634  
 |||||  
 1851 AAAAGTGGTTTATCAAGGCAATATCATCGCCCATTTATCAAAAAATCAG 1900  
 |||||  
 634 aLeuArgGlyTyrAlaLysLeuGlyTyrGlyAsnAsnLeuProPheTyr 650  
 |||||  
 1901 TCTTCGGTGGATACCCCAAGTAGCTATGGCAATTAATTTACCATTTAT 1950  
 |||||  
 651 GluAsnPheTyrAlaGlyTyrGlySerValArgGlyTyrAspGlnse 667  
 |||||  
 1951 GAAAAATTTCTATGACGCGCGGTATGTTGCGTTCGTGGCTATGATCAATC 2000  
 |||||  
 667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684  
 |||||  
 2001 CTCCTTGGGTCACCTCACAAGCCATATTGACAGCTCGTCGTGTCAC 2050  
 |||||  
 684 InThrThrLeuGlyGluValValGlyGlyAsnAlaLeuAlaThrPheGly 700  
 |||||  
 2051 AAACACACTAGGACAGGTGTTGGTGTATGCTTTGGCAACTTTCGGC 2100  
 |||||  
 701 SerGluLeuIleLeuProLeuProPheLysGlyAspTrpIleAspGlnVa 717  
 |||||  
 2101 ACTGAGCTGATTTTACCTTTGCGCATTTAAAGGTGATTGATAGATCAGT 2150  
 |||||  
 717 lArgProValIlePheIleGluGlyGlyGlnValPheAspThrThrGlyM 734  
 |||||  
 2151 GCGTCCAGTGATTTATTTGAGGGCGGTGAGGTTTGTATACAAACAGSTA 2200  
 |||||  
 734 eAspLysGlnThrIleAspLeuThrGlnPheLysAspProGlnAlaThr 750  
 |||||  
 2201 TGGATAACAACCACTGATTGTTAAACCAATTTAAAGACCCCAACGACACA 2250  
 |||||  
 751 AlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuThrGlnAspLy 767  
 |||||  
 2251 GCTGAACAAAATGCAAAAGCAGCCCAATCGCCGCTACTAACCAAGATAA 2300  
 |||||  
 767 scGlnLeuArgTyrSerAlaGlyValGlyAlaThrTrpTyrThrProIleG 784  
 |||||  
 2301 ACAGTTGCGGTATAGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2350  
 |||||  
 784 lyProLeuSerIleSerTyrAlaLysProLeuAsnLysLysGlnAsnAsp 800  
 |||||  
 2351 GTCCTTTATCTATTAGCTATGCCAGCCATTGAATAAAAAACAATAATGAT 2400  
 |||||  
 801 GlnThrAspThrValGlnPheGlnIleGlySerValPhe 813  
 |||||  
 2401 CAGACCGGATACGGTACAGTTCAGATTGGTAGTGTCTTT 2439  
 |||||

seq\_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AAZ29551

seq\_documentation\_block:

ID AAZ29551 standard; DNA; 2442 BP.

XX AC AAZ29551;

XX AC AAZ29551;

XX AC AAZ29551;

XX AC AAZ29551;

XX AC AAZ29551;

XX AC AAZ29551;

XX AC AAZ29551;

XX AC AAZ29551;

XX AC AAZ29551;

XX AC AAZ29551;

XX AC AAZ29551;

XX AC AAZ29551;

XX AC AAZ29551;

XX AC AAZ29551;

XX AC AAZ29551;

XX AC AAZ29551;

XX AC AAZ29551;

PN WO9963093-A2.  
 XX 09-DEC-1999.  
 XX 31-MAY-1999; 99WO-EP03822.  
 PF 03-JUN-1998; 98GB-0011945.  
 PR 08-MAR-1999; 99GB-0005304.  
 XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 XX Vinals-Bassols C;  
 PI WPI; 2000-105700/09.  
 DR P-PSDB; AAY44391.  
 XX Novel BASB027 polynucleotide and polypeptides from Moraxella  
 PT catarrhalis useful for treating M. catarrhalis infection such as otitis  
 PT media  
 XX Claim 8; Page 102; 109pp; English.

XX The present sequence is a DNA obtained from chromosomal DNA library of  
 CC Moraxella catarrhalis strain Mc2931 (ATCC 43617). It encodes BASB027  
 CC polypeptide, which shows significant homology to Neisseria meningitidis  
 CC OMP85 outer membrane protein. BASB027 polynucleotide and polypeptide can  
 CC be used for diagnosis and staging of disease, determining susceptibility  
 CC to a disease and to prepare medicaments for treating M. catarrhalis  
 CC infections, especially otitis media. The BASB027 DNA can be used as  
 CC probe for screening of genetic mutations, serotype, taxonomic  
 CC classification or identification. BASB027 agonists, antagonists and  
 CC antibodies may be used to prevent and/or treat bacterial infections.

XX Sequence 2442 BP; 735 A; 461 C; 535 G; 711 T; 0 other;

alignment\_scores:

Quality: 749.00 Length: 749

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-701-711-2 x AAZ29551

Align seg 1/1 to: AAZ29551 from: 1 to: 2442

65 GlnLeuAlaAspGlyValLysAlaLeuTyrAlaThrGlyAsnPheserAs 81

|||||

193 CAGTTGGCTGATGGTGTCAAGCACTTTATGCAACAGGCAATTTTTCAGA 242

|||||

81 pValGlnValTyrHisGlnGluGlyArgIleTyrGlnValThrGluA 98

|||||

243 TGTGCAAGTCTATCATCAAGAAGGGGCTATCATCTATCAGTAACCGAAA 292

|||||

98 rgProLeuIleAlaGluIleAsnPhesGluGlyAsnArgLeuIleProLys 114

|||||

293 GSCCGCTAATCGCTGAGATTAATTTTGGGCAATCGCTTAATTCACAAA 342

|||||

115 GluGlyLeuGlnGluGlyLeuLysAsnAlaGlyLeuAlaValGlyGlnPr 131

|||||

343 GAAGGCTACAAAGAGGGCTAAAAAATGCTGGCTTACGTTGGGTCAACC 392

|||||

131 oLeuLysGlnAlaThrValGlnMetIleGluThrGluLeuThrAsnGlnT 148

|||||

393 ACTAAAAAGCCACAGTACATGATCGAATCGAATTAATCTCAACAGAGATG 442

|||||

148 yTleSerGlnGlyTyrTyrAsnThrGluIleThrValLysGlnThrMet 164

|||||

443 ATATATCACAGGCTATTATATACCGAATTAATCTCAACAGAGATG 492

|||||

165 LeuAspGlyAsnArgValLysLeuAspMetThrPheAlaGluGlyLysPr 181

|||||

493 CTTGATGGTAATCGTGTAAAGCTTGATATGACCTTTGCTGAAAGGTAACC 542

181 oAlaArgValValAspIleAsnIleLeGlyAsnGlnHisPheSerAspA 198  
|||||  
543 TGCACGGGTGGTTCATATTATATCATTCATGCAATCAGCATTTAGCGATG 592  
198 laAspLeuIleAspValLeuAlaIleLysAspAsnLysIleAsnProLeu 214  
|||||  
593 CAGATTTGATGTGTTGCGTAAAGGATAATAAAATCAATCCACTG 642  
215 SerLysAlaAspArgTyrThrGlnGluLysLeuValThrSerLeuGluAs 231  
|||||  
643 TCATAAGCTCAGCGTTATATCTCAGAAAGCTGGTACCAGTTTAGAGAA 692  
231 nLeuArgAlaLysTyrLeuAsnAlaGlyPheValArgPheGluIleLysA 248  
|||||  
693 TTTGCGGTGCTAAATATCTCAATCAGCGTTGTGCGTTTTGAGATTAAG 742  
248 spAlaLysLeuAsnIleAsnGluAspLysAsnArgIlePheValGluIle 264  
|||||  
743 ATGCTAAGCTTAATATTAATGAAGATAAAACCGTATCTTTGTTGAGATT 792  
265 SerLeuHisGluGlyGlnTyrArgPheGlyGlnThrGlnPheLeuG1 281  
|||||  
793 TCATTGCTGAAGGTGAGCAATATCGCTTTGGACAGACAGATTTTGGG 842  
281 yAsnLeuThrTyrThrGlnAlaGluLeuGluAlaLeuLeuLysPheLysA 298  
|||||  
843 TAATTTAACTTATACTCAACGCAACTTGAGGCACATGCTTAATTCAAAG 892  
298 laGluGluGlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsnIle 314  
|||||  
893 CAGAAGAAGGGTTTTCACAAAGCCATGCTTGAGCAAAACAAACAATATC 942  
315 SerThrLysPheGlyAspAspGlyTyrTyrTyrAlaGlnIleArgProVa 331  
|||||  
943 AGTACCATAATTTGGTGACGAGTATATATGCTCAAAATCCGCTCTGT 992  
331 lThrArgIleAsnAspGluSerArgThrValAspValGluTyrTyrIleA 348  
|||||  
993 AACACGCATTAATGATCAAAAGTCGTACGGTTGATGTGGAATATATATTG 1042  
348 spProValHisProValTyrValArgArgIleAsnPheThrGlyAsnPhe 364  
|||||  
1043 ACCCTGTACACCTGCTCTATGTACGGCGTATTAATTTTACAGTAACCTT 1092  
365 LysThrGlnAspGluValLeuArgArgGluMetArgGlnLeuGluGlyAl 381  
|||||  
1093 AAGACCCAAGATCAAGTACTCCCGTCGTGAGATCGCAACACTTGAAGTGC 1142  
381 aLeuAlaSerAsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArgT 398  
|||||  
1143 GTTGGCATCTAATCAAAAAATCCAGCTGCTCGTGCACGCTTGATCGGA 1192  
398 hrGlyPhePheLysHisValThrValAspThrArgProValProAsnSer 414  
|||||  
1193 CTGGGTTTTTAAACATGTTTACCGTTGATCTCGTCCAGTACCACCACTCA 1242  
415 ProAspGlnValAspValAsnPheValValGluGluGlnProSerGlySe 431  
|||||  
1243 CCTGATCAGGTTCATGTAATTTTGTGTTGAAGACACACCTTCAGGATC 1292  
431 rSerThrIleAlaAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnP 448  
|||||  
1293 ATCAACCATCGCAGCAGCGGTACTCTCAAGTGGTGTGAACCTTTTCAAT 1342  
448 heAspValSerGlnAsnAsnPheMetGlyThrGlyLysHisValAsnAla 464  
|||||  
1343 TTGATGTTCTCAAAATACCTTTATGGGTACAGGTAAACGACGTCATGCT 1392  
465 SerPheSerArgSerGluThrArgGluValTyrSerLeuGlyMetThrAs 481  
|||||  
1393 TCGTTTTCTCGCTCTCAGACCCCGTGAAGTGTATAGTTTGGGTATGACCAA 1442

481 nProTyrPheThrValAsnGlyValSerGlnSerLeuSerGlyTyrTyrA 498  
|||||  
1443 CCCATACTTTTACCCTAAATGGCGTCTCGCAAGCTTGAGTGGCTACTATC 1492  
498 rgLysThrLysTyrAspAsnLysAsnIleSerAsnTyrValLeuAspSer 514  
|||||  
1493 GTAAACCAAGTATGATAACAAGAACATTAGTAATTATGACTGATTCT 1542  
515 TyrGlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgI1 531  
|||||  
1543 TATGGTGGCTCATTAAGCTATGGATATCCAAATTCATGATAAATCAACGAT 1592  
531 eSerPheGlyLeuAsnAlaAspAsnThrLysLeuHisGlyGlyArgPheM 548  
|||||  
1593 AAGCTTTGGTCTGAATGCTCACAATACCAAGCTTCATGGCGGTGCTTTTA 1642  
548 etGlyIleSerAsnValLysGlnLeuMetAlaaspGlyGlyLysIleGln 564  
|||||  
1643 TGGGCATTTAGTAATGTCAAGCAGCTGATGGCAGATGGTGCAAAATTCAA 1692  
565 ValAspAsnAsnGlyIleProAspPheLysHisAspTyrThrTyrAs 581  
|||||  
1693 GTGGATAATAATGGCATTCCTGATTTTAAACATGATTACACACCTACAA 1742  
581 nAlaIleLeuGlyTyrPAsnTyrSerSerLeuAspArgProValPheProt 598  
|||||  
1743 TGCCATTTTGGGTGGAATTTATCAAGCTAGATCGCCCTGTATTTCCAA 1792  
598 hrGlnGlyMetSerHisSerValAspLeuThrValGlyPheGlyAspLys 614  
|||||  
1793 CCCAAGGCATGAGTCATCTCTAGATTTTGACGGTTGGTTTGGTGATAAA 1842  
615 ThrHisGlnLysValValTyrGlnGlyAsnIleTyrArgProPheIleLy 631  
|||||  
1843 ACTCATCAAAAAGTGGTTTATCAAGGCATATCTATGCCCCCTTTATCAA 1892  
631 sLysSerValLeuArgGlyTyrAlaLysLeuGlyTyrGlyAsnAsnLeup 648  
|||||  
1893 AAAATCAGTCTTGGTGGATACGCCAAGTTAGGCTATGGCAATAATTTAC 1942  
648 roPheTyrGluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyr 664  
|||||  
1943 CATTTATGAAAATTTCTATGACAGGCGGTATGGTTCGGTTCGGTGTAT 1992  
665 AspGlnSerSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgAr 681  
|||||  
1993 GATCAATCTCTTTGGTCCACGCTCACAAAGCTATTTGACAGCTCGTCG 2042  
681 gGlyGlnGlnThrThrLeuGlyGluValValGlyGlyAsnAlaLeuAlat 698  
|||||  
2043 TGGTCAACAAACCACACTAGGAGAGGTTGTTGGTGAATGCTTTGGCAA 2092  
698 hrPheGlySerGluLeuIleLeuProLeuProPheLysGlyAspTrrPle 714  
|||||  
2093 CTTTCCGAGTGAGCTGATTTTACCTTTTGCCATTTAAAGGTGATGGATA 2142  
715 AspGlnValArgProValIlePheIleGluGlyGlyGlnValPheAspTh 731  
|||||  
2143 GATCAGGTGGTCCAGTGATATTCATTTAGGCGCGTCAGCTTTTGTATAC 2192  
731 rThrGlyMetAspLysGlnThrIleAspLeuThrGlnPheLysAspProG 748  
|||||  
2193 AACAGGTATGGATAAAACAAACCATTTGATTTTAAACCAATTTTAAAGACCCAC 2242  
748 InAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuThr 764  
|||||  
2243 AAGCAACAGCTGAACAAATGCAAAAGCAGCCCAATCGCCCGCTACTRACC 2292  
765 GlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrTrpTyrTh 781  
|||||  
2293 CAAGATAAACAGTTCGTTATAGTGTGGTGGTGGTGGTGGTGGTGGTGGT 2342  
781 rProIleGlyProLeuSerIleSerTyrAlaLysProLeuAsnLysLysG 798



|||||  
2343 GCCCATTGGTCCTTTATCTATTAGCTATGCCAAGCCATTGAATAAAAAAC 2392

798 InAsnAspGlnThrAspThrValGlnPheGlnIleGlySerValPhe 813  
|||||

2393 AAATGATGACAGCGATACGATCCAGATTGGTAGTGCTTT 2439  
|||||

seq\_name: /SDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AAF28550

seq\_documentation\_block:

ID AAF28550 standard; DNA; 99629 BP.

XX AAF28550;

XX 04-APR-2001 (first entry)

XX Genomic fragment #37.

XX Genomic library; bacteria; human upper airway; otitis media; sinusitis;  
KW bronchopulmonary; endocarditis; meningitis; ss.

XX Moraxella catarrhalis.

XX WO200078968-A2.

XX 28-DEC-2000.

XX 16-JUN-2000; 2000WO-US16649.

XX 18-JUN-1999; 99US-0140121.

XX (INCY-) INCYTE GENOMICS INC.

XX Lagace RE, Patterson C, Berg KL;

XX WPI; 2001-041427/05.

XX Genomic library for identifying diagnostic and therapeutic  
PT compositions, and for identifying virulence factors, regulatory  
PT elements and drug targets, comprises Moraxella catarrhalis nucleic  
PT acids -

XX Claim 1; Page 391-415; 545pp; English.

XX The present invention relates to a Moraxella catarrhalis genomic library  
CC comprising a combination of 41 nucleic acid molecules (see  
CC AAF28514-AAF28554). The library has a number of uses described in the  
CC specification e.g. is useful for identifying diagnostic and therapeutic  
CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large  
CC aerobic, gram-negative diplococcus, normally found among the bacterial  
CC flora of human upper airways. M. catarrhalis is known to cause acute,  
CC localised infections such as otitis media, sinusitis and bronchopulmonary  
CC infection and life-threatening, systemic diseases including endocarditis  
CC and meningitis.

XX SQ Sequence 99629 BP; 29233 A; 19222 C; 21909 G; 29264 T; 1 other;

alignment\_scores:

Quality: 749.00 Length: 749  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-701-711-2 x AAF28550 ..

Align seg 1/1 to: AAF28550 from: 1 to: 99629

65 GlnLeuAlaAspGlyValLysAlaLeuTyrAlaThrGlyAsnPheSerAs 81  
|||||

66325 CAGTTGGCTGATGGTCAAGCACTTTATGCAACAGCGCAATTTTCAGA 66374  
|||||

81 pValGlnValTyrHisGlnGluGlyArgIleIleTyrGlnValThrGluA 98  
|||||

|||||  
66375 TGTCCAAGTCTCATCATCAAGAAGGCGCTATCATCTATCAGGTAACCGAAA 66424  
|||||

98 rgProLeuIleAlaGlnIleAsnPheGluGlyAsnArgLeuIleProLys 114  
|||||

66425 GGCCGTTAATCCCTGAGATTAAATTTTGGAGGCAATCGCTTAATTCACAAA 66474  
|||||

115 GluGlyLeuGlnGluGlyLeuLysAsnAlaGlyLeuAlaValGlyGlnPr 131  
|||||

66475 GAAGGTCTACAAGAAGGCGTAAAAAATGCTGGCTTAGCTGGGTCAACC 66524  
|||||

131 oLeuLysGlnAlaThrValGlnMetIleGluThrGluLeuThrAsnGlnT 148  
|||||

66525 ACTAAACCAAGCCACAGTACAGATCGAAACCCAGCTTACCAATCAAT 66574  
|||||

148 yrIleSerGlnGlyTyrTyrAsnThrGluIleThrValLysGlnThrMet 164  
|||||

66575 ATATATCACAGCGTATTATTAATACCGAAATTAAGTCAACACAGCATG 66624  
|||||

165 LeuAspGlyAsnArgValLysLeuAspMetThrPheAlaGluGlyLysPr 181  
|||||

66625 CTTGATGGTAATCGTTAAGCTTGATATGACCTTTGCTGAAGGTAAACC 66674  
|||||

181 oAlaArgValValAspIleAsnIleIleGlyAsnGlnHisPheSerAspA 198  
|||||

66675 TGCACGGGTGGTTGATATTAATATCATTTGGCAATCAGCATTTTAGCGATG 66724  
|||||

198 laAspLeuIleAspValLeuAlaIleLysAspAsnLysIleAsnProLeu 214  
|||||

66725 CAGATTTGATTGATGCTTCGGATTAAAGGATATAAAATCAATCCACTG 66774  
|||||

215 SerLysAlaAspArgTyrThrGlnGluLysLeuValThrSerLeuGluAs 231  
|||||

66775 TCCTAAAGCTGACCGTTATACTCAAGAAAAGCTGGTGACCACTTTAGACAA 66824  
|||||

231 nLeuArgAlaLysTyrLeuAsnAlaGlyPheValArgPheGluIleLysA 248  
|||||

66825 TTTGCGTGCTAAATATCTCAATGCAAGGTTTGTGCGTTTGGAGATTAAAG 66874  
|||||

248 spAlaLysLeuAsnIleAsnGluAspLysAsnArgIlePheValGluIle 264  
|||||

66875 ATGCTAAGCTTAATTAATGAAGATAAAACCGTATCTTTGTTGAGATT 66924  
|||||

265 SerLeuHisGluGlyGluGlnTyrArgPheGlyGlnThrGlnPheLeuGl 281  
|||||

66925 TCATTGTCATGAAGTGAGCATATCGCTTTGGACAGACACAGATTTTGGG 66974  
|||||

281 yAsnLeuThrTyrThrGlnAlaGluLeuGluAlaLeuLysPheLysA 298  
|||||

66975 TAATTTAACTTATACTCAAGCAGAACTTGAGGCACTGCTTAAATTCAAA 67024  
|||||

298 laGluGluGlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsnIle 314  
|||||

67025 CAGAAGAAGGGTTTTCACAGCCATGCTTGAGCAAAACAAACAATATC 67074  
|||||

315 SerThrLysPheGlyAspAspGlyTyrTyrTyrAlaGlnIleArgProVa 331  
|||||

67075 AGTACCAAAATTTGGTGAGGATGGCTATTATTATGCTCAAAATCCGCTCTGT 67124  
|||||

331 lThrArgIleAsnAspGluSerArgThrValAspValGluTyrTyrIleA 348  
|||||

67125 AACACGCATTAAATGATGAAGTCGTACGGTTGATCTGGAATATTATATTG 67174  
|||||

348 spProValHisProValTyrValArgArgIleAsnPheThrGlyAsnPhe 364  
|||||

67175 ACCCTGTACACCCCTGTATGTACCGCGTATTAATTTTACAGGTAACCTT 67224  
|||||

365 LysThrGlnAspGluValLeuArgArgGluMetArgGlnLeuGluGlyAl 381  
|||||

67225 AAGACCCAGATGAAGTACTCCGTCGTGAGATGCCACAACCTTGAAGGTGC 67274  
|||||

381 aLeuAlaSerAsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArgT 398  
|||||

67275 GTTGGCATCTAATCAAAAAATCAGCTGCTCGTCACACGCTTGATGCGGA 67324  
398 hrGlyPhePheLysHisValThrValAspThrArgProValProAsnSer 414  
67325 CTGGGTTTTTAAACAATGTTACCGTTGATCTCGCTGATACCCAACTCA 67374  
415 ProAspGlnValAspValAsnPheValValGluGlnProSerGlyse 431  
67375 CCTGATCAGGTTGATGATAAATTTGGTGTGAAGAACACCTTCAGGATC 67424  
431 rSerThrIleAlaAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnP 448  
67425 ATCAACCATCGCAGCAGGCTACTCTCAAAAGTGGTGGTGAACCTTTTCAAT 67474  
448 heAspValSerGlnAsnAsnPheMetGlyThrGlyLysHisValAsnAla 464  
67475 TTGATGTTTCTCAAAATAACTTTATGGGTACAGGTAAAGCAGCTCAATGCT 67524  
465 SerPheSerArgSerGlnThrArgGluValTyrSerLeuGlyMetThrAs 481  
67525 TCGTTTTCCTCGCTCTGAGACCCGCTGAGGTGTATAGTTTGGGTATGACCAA 67574  
481 nProTyrPheThrValAsnGlyValSerGlnSerLeuSerGlyTyrTyra 498  
67575 CCATACCTTTACCGTAAATGGCGCTCGCAAAAGCTTGAGTGGCTACTATC 67624  
498 rGlySerThrLysTyrAspAsnLysAsnIleSerAsnTyrValLeuAspSer 514  
67625 GTAAACCAAGTATGATAACAAGACATTTAGTAATATGCTACTTGATTCT 67674  
515 TyrGlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIle 531  
67675 TATGTTGGCTCATTAAGCTATGATATCAATTCATGATGAAATCAACGCAT 67724  
531 eSerPheGlyLeuAsnAlaAspAsnThrLysLeuHisGlyGlyArgPheM 548  
67725 AGCTTTGCTGTAATGCTGCAATACCAAGCTTCAATGGCGGTGCTTTTA 67774  
548 etGlyIleSerAsnValLysGlnLeuMetAlaAspGlyGlyLysIleGln 564  
67775 TGGGCATTTAGTAATGTCAGCAGCTGATGCCAGATGTTGGCAAAATTCAA 67824  
565 ValAspAsnAsnGlyIleProAspPheLysHisAspTyrThrTyrAs 581  
67825 GTGGATAATAATGGCATCTCCTGATTTAAGCATGATTACACAACTTACAA 67874  
581 nAlaIleLeuGlyTyrAsnTyrSerSerLeuAspArgProValPheProT 598  
67875 TGCCATTTTGGGTGGATATTCAGTCTAGATCGCCCTGTATTTCCAA 67924  
598 hrGlnGlyMetSerHisSerValAspLeuThrValGlyPheGlyAspLys 614  
67925 CCCAAGCATGATGATCTATTCTGATAGATTTGACGTTGGTTTGGTGATAAA 67974  
615 ThrHisGlnLysValValTyrGlnGlyAsnIleTyrArgProPheIleIly 631  
67975 ACTCATCAAAAAGTGGTTTATCAAGCAATATCTATCGCCCATTTATCAA 68024  
631 sLysSerValLeuArgGlyTyrAlaLysLeuGlyTyrGlyAsnAsnLeuP 648  
68025 AAAATCAGTCTTGGGTGGATAGCCCAAGTAGGCTATGGCAATAATTTAC 68074  
648 roPheTyrGluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyr 664  
68075 CATTTTATCAAAATTTCTATGACGGCGGTATGTTGCTGGTTCGTCGTAT 68124  
665 AspGlnSerSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgAr 681  
68125 GATCAATCTCTTTGGTCCAGCTCACAAAGCTTATTTGACAGCTCGTGC 68174  
681 gGlyGlnGlnThrThrLeuGlyGlnValValGlyGlyAsnAlaLeuAlat 698  
68175 TGGTCAACAAACACACTAGAGAGAGTGTGGTGGTAATGCTTTGGCAA 68224

698 hrPheGlySerGluLeuIleLeuProLeuProPheLysGlyAspTrpIle 714  
68225 CTTTCGGCAGTCACTGATTTTACCTTTGCCATTTAAAGTGATTCGATA 68274  
715 AspGlnValArgProValIlePheIleGluGlyGlyGlnValPheAspTh 731  
68275 GATCAGTCCGTCAGTGATATTCAATTGAGGCGGTGAGTTTTCATAC 68324  
731 rThrGlyMetAspLysGlnThrIleAspLeuThrGlnPheLysAspProG 748  
68325 AACAGGTATGGATAACAACCACTGATTAAACCCCAATTTAAAGACCCAC 68374  
748 InAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuThr 764  
68375 AAGCAACAGCTGAACAAATGCAAAAGCAGCAATCGCCGCTACTAAC 68424  
765 GlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrTrpTyrTh 781  
68425 CAAGATAAACAGTTGCGTTATAGTGTGTTGGTGCAACTTGGTATAC 68474  
781 rProIleGlyProLeuSerIleSerTyrAlaLysProLeuAsnLysLysG 798  
68475 GCCCATGTGCTCTTATCTATTAGCTATGCCAAGCCRTTGAATATAAAAC 68524  
798 InAsnAspGlnThrAspThrValGlnPheGlnIleGlySerValPhe 813  
68525 AAAATGATCAGACCGATACGTTACAGTTCAGATTGGTAGTGCTTT 68571  
seq\_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AAF91415  
seq\_documentation\_block:  
ID AAF91415 standard; DNA; 1000 BP.  
XX  
AC AAF91415;  
XX  
DT 04-MAY-2001 (first entry)  
XX  
DE Moraxella catarrhalis D15 gene upstream sequence, SEQ ID:41.  
XX  
KW Modified Gram-negative bacterium; outer membrane vesicle; bleb; vaccine;  
KW genetically modified; protective antigen expression; LPS detoxification;  
KW LPS; lipid A; homologous recombination vector; immunisation;  
KW immunoprotective; non-toxic; paediatric; ds.  
XX  
OS Moraxella catarrhalis.  
XX  
PN WO200109350-A2.  
XX  
PD 08-FEB-2001.  
XX  
PF 31-JUL-2000; 2000WO-EP07424.  
XX  
PR 03-AUG-1999; 99GB-0018319.  
XX  
PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
XX  
PI Berthet FJ, Dalemans WLJ, Denoel P, Dequesne G, Feron C, Lobet Y;  
PI Poolman J, Thily G, Thonnard J, Voet P;  
XX  
XX WPI; 2001-138654/14.  
XX  
PT New isolated polynucleotide useful for outer membrane vesicle  
PT preparation from Gram-negative bacterial strain for vaccination of  
PT microbial infections .  
XX  
PS Claim 46; Page 87; 128pp; English.  
XX  
CC The invention relates to a genetically-engineered outer membrane vesicle  
CC (bleb) preparation from a Gram-negative bacterium for use as a vaccine.  
CC The blebs of the invention are improved with respect to their  
CC immunogenicity and toxicity by the introduction of one or more genetic  
CC changes to the chromosome of the bacterium from which the blebs are

CC derived. The changes made include the upregulation of protective antigen  
CC expression, the downregulation of immunodominant non-protective antigen  
CC expression, and genetic changes which result in detoxification of the  
CC Lipid A moiety of lipopolysaccharide (LPS). The invention also  
CC encompasses modified gram-negative bacterial strains from which the bleb  
CC preparations are made, a vector suitable for performing recombination  
CC events (for the generation of the modified bacterial strains),  
CC bacterially-derived nucleic acid sequences used in such a vector, and an  
CC immunoprotective and non-toxic Gram-negative bleb, ghost, or killed whole  
CC cell vaccine suitable for paediatric use. The bleb preparation is useful  
CC in the manufacture of a medicament for immunising a human host against a  
CC disease caused by infection of one or more of the following: *Neisseria*  
CC meningitidis, *Neisseria gonorrhoeae*, *Haemophilus influenza*, *Moraxella*  
CC catarrhalis, *Pseudomonas aeruginosa*, *Chlamydia trachomatis*, and *Chlamydia*  
CC pneumoniae. The invention may also be used to provide immunisation against  
CC the influenza virus. Bacterially derived nucleotide sequences of the  
CC invention are used in the performance of homologous recombination events  
CC up to 1000 bp upstream of a bacterial chromosomal gene in order to either  
CC increase or decrease expression of that gene. Immunoprotective and  
CC non-toxic Gram-negative bleb, ghost, or killed whole cell vaccines  
CC are more immunogenic, less toxic and safer, and are particularly useful  
CC for paediatric use. The present sequence represents a specifically  
CC claimed *Moraxella catarrhalis* nucleic acid sequence.  
XX  
SQ Sequence 1000 BP; 287 A; 172 C; 215 G; 326 T; 0 other;

alignment\_scores:  
Quality: 36.00 Length: 36  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-701-711-2 x AAF91415 ..

Align seg 1/1 to: AAF91415 from: 1 to: 1000

1 MetArgAsnSerTyrPheLysGlyPheGlnValSerAlaMetThrMetAla 17  
|||||  
821 ATGGCTAATTCATATTTAAAGGTTTTCAGGTCAGTGCATGACAATGGC 870  
  
17 aValMetMetValMetSerThrHisAlaGlnAlaAlaAspPheMetAla 34  
|||||  
871 TGTGATGATGGTAATGTCAACTCATGTCACAAGCGCGGATTATTATGGCAA 920

34 snAspIle 36  
|||||

921 ATGACATT 928

seq\_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAZ29554

## seq\_documentation\_block:

ID AAZ29554 standard; DNA; 61 BP.

AC AAZ29554;

DT 14-MAR-2000 (first entry)

DE M. catarrhalis BASB027 amplifying forward PCR primer MC-D15-BamF.

XX PCR primer; MC-D15-BamF; BASB027; otitis media; OMP85;

KW outer membrane protein; treatment; diagnosis; bacterial infection; ss.

XX Synthetic.

OS *Moraxella catarrhalis*.

XX WO963093-A2.

PN 09-DEC-1999.

XX 31-MAY-1999; 99WO-EP03822.

XX 03-JUN-1998; 98GB-0011945.

XX

PR 08-MAR-1999; 99GB-0005304.  
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
PA Vinals-Bassols C;  
XX WPI; 2000-105700/09.  
DR Novel BASB027 polynucleotide and polypeptides from *Moraxella*  
XX catarrhalis useful for treating *M. catarrhalis* infection such as otitis  
PT media -  
XX Example 2; Page 51; 109pp; English.  
XX The present sequence is a forward PCR primer MC-D15-BamF used for  
CC amplifying chromosomal DNA extracted from 16 *Moraxella catarrhalis*  
CC strains for variably analysis of the BASB027 gene.  
XX Sequence 61 BP; 20 A; 15 C; 17 G; 9 T; 0 other;

## alignment\_scores:

Quality: 11.00 Length: 11  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-701-711-2 x AAZ29554 ..

Align seg 1/1 to: AAZ29554 from: 1 to: 61

39 ThrGlyLeuGlnArgValThrIleGluSerLeu 49  
|||||  
28 ACAGGACTACAGCGAGTGACCATTCAGAAAGCTTA 60

seq\_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAZ29555

## seq\_documentation\_block:

ID AAZ29555 standard; DNA; 67 BP.

AC AAZ29555;

DT 14-MAR-2000 (first entry)

DE M. catarrhalis BASB027 amplifying reverse PCR primer MC-D15-SalRC.

XX PCR primer; MC-D15-SalRC; BASB027; otitis media; OMP85;

KW outer membrane protein; detection; treatment; bacterial infection; ss.

XX Synthetic.

OS *Moraxella catarrhalis*.

XX WO963093-A2.

XX 09-DEC-1999.

XX 31-MAY-1999; 99WO-EP03822.

XX 03-JUN-1998; 98GB-0011945.

XX 08-MAR-1999; 99GB-0005304.

XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.

XX Vinals-Bassols C;

XX WPI; 2000-105700/09.

XX Novel BASB027 polynucleotide and polypeptides from *Moraxella*  
PT catarrhalis useful for treating *M. catarrhalis* infection such as otitis  
PT media -  
XX Example 2; Page 51; 109pp; English.

XX

XX

CC The present sequence is a reverse complementary PCR primer MC-D15-SalRC  
 CC used for amplifying chromosomal DNA extracted from 16 Moraxella  
 CC catarrhalis strains for variability analysis of the BASB027 gene.  
 XX  
 SQ Sequence 67 BP; 22 A; 16 C; 15 G; 14 T; 0 other;

alignment\_scores:  
 Quality: 11.00 Length: 11  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-701-711-2 x AAZ29555/rev ..  
 Align seg 1/1 to reverse of: AAZ29555 from: 1 to: 67

803 AspThrValGlnPheGlnIleGlySerValPhe 813  
 |||||  
 66 GATACGTACAGTCCAGATGAGTGTCTTT 34

seq\_name: /SIDS1/gcgdata/geneseq/geneseq-n-emb1/NA2001A.DAT.AAF60966

seq\_documentation\_block:  
 ID AAF60966 standard; DNA; 1830 BP.

XX AAF60966;

XX 16-MAY-2001 (first entry)

DE P. putida KT2440-associated DNA ORF00612.

XX Transgenic plant; detection; probe; amplification; vaccine carrier;  
 KW microbial production strain; biological remediation; ds.

OS Pseudomonas putida.

XX DE19935088-A1.

PD 01-FEB-2001.

XX 27-JUL-1999; 99DE-1035088.

XX 27-JUL-1999; 99DE-1035088.

XX (TIGR-) TIGR INST GENOMIC RES.

PA (QUITA-) QUIAGEN GMBH.

PA (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.

PA (DKFZ-) DKFZ DEUT KREBSFORSCHUNGSZENTRUM.

PA (MEDI-) MEDIZINISCHE HOCHSCHULE HANNOVER.

XX WPI; 2001-192469/20.

XX New DNA sequences specific for Pseudomonas putida KT2440, useful as  
 PT safe genetic engineering host, allow detection in presence of other  
 PT related bacteria -

XX Claim 1a; Page 13-14; 158pp; German.

XX This invention describes novel DNA sequences (I) for specific detection  
 CC of Pseudomonas putida KT2440. The invention also describes (1)  
 CC recombinant expression vector containing (1); (2) prokaryotic or  
 CC eukaryotic cells transformed or transfected with (1) or the vector of  
 CC (1); (3) production of expression products by culturing cells of (2);  
 CC (4) expression products, or their fragments, of (1) and synthetic  
 CC proteins or peptides with the same sequences (A); (5) poly- or  
 CC mono-clonal antibodies (Ab) that react specifically with (A); (6)  
 CC hybridoma cells that produce the monoclonal Ab of (5); (7) transgenic  
 CC plants that contain transformed or transfected cells of (2); (8)  
 CC detecting KT2440 using a labeled (1) or Ab as probe; and (9) DNA chips  
 CC carrying one or more (1). (1), and their fragments, are used as probes  
 CC to detect and isolate full-length cDNAs and/or to amplify such cDNAs by  
 CC polymerase chain reaction, and for production of transgenic plants. (1),

CC or antibodies that recognize their expression products, are used for  
 CC detecting the presence of KT2440, particularly in presence of other,  
 CC even closely related, bacteria. KT2440 is one of the bacteria classified  
 CC as safe, by the National Institutes of Health, for genetic engineering  
 CC work, e.g. as microbial production strains, for biological remediation  
 CC and as vaccine carriers. (1) are exclusive to KT2440 with no significant  
 CC homology with sequences in other bacteria (specifically the closely  
 CC related pathogen P. aeruginosa). Compared with other 'safe' bacteria, it  
 CC has greater catabolic activity and better survival in, and adaptation to,  
 CC the rhizosphere and soil.  
 XX  
 SQ Sequence 1830 BP; 401 A; 555 C; 529 G; 344 T; 1 other;

alignment\_scores:  
 Quality: 10.00 Length: 10  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-701-711-2 x AAF60966 ..

Align seg 1/1 to: AAF60966 from: 1 to: 1830

368 AspGluValLeuArgArgGluMetArgGln 377  
 |||||  
 1081 GACGAAGTGTGCGTCCGAAATCGCCAG 1110

seq\_name: /SIDS1/gcgdata/geneseq/geneseq-n-emb1/NA2001B.DAT.AAS85853

seq\_documentation\_block:  
 ID AAS85853 standard; cDNA; 3750 BP.

XX AAS85853;

XX 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #21657.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX P-PSDB; ABG21666.

XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -

PS Claim 1; SEQ ID No 21657; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 3750 BP; 937 A; 928 C; 965 G; 920 T; 0 other;

alignment\_scores:  
 Quality: 10.00 Length: 10  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-701-711-2 x AAS85853 ..

Align seg 1/1 to: AAS85853 from: 1 to: 3750

648 ProPheTyrGluAsnPhetYrAlaGlyGly 657  
 |||||  
 1435 CGGTTCTACGAGACTTATGCCGTGGT 1464

seq\_name: /SIDSL/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:AAT84180

seq\_documentation\_block:

ID AAT84180 standard; DNA; 1000 BP.

XX AC AAT84180;

XX DT 14-SEP-1998 (first entry)

XX DE DNA encoding a Staphylococcus aureus protein of unknown function.

XX Staphylococcus aureus protein; ribozyme; antisense sequence; control;  
 KW Staphylococcal gene; regulatory element; bacterial gene expression;  
 KW vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;  
 KW toxic shock syndrome; ss.

XX OS Staphylococcus aureus.

XX Key Location/Qualifiers  
 FT CDS complement (357..647)  
 FT /\*tag= a

XX PN WO9730070-A1.

XX PD 21-AUG-1997.

XX PF 19-FEB-1997; 97WO-US02318.

XX PR 20-FEB-1996; 96US-0011888.

XX (SMIK ) SMITHKLINE BEECHAM CORP.

XX Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO;  
 PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;

XX WPI; 1997-424969/39.

XX P-PSDB; AAW28281.

XX Novel polypeptide(s) from Staphylococcus aureus strains WCUH29 - used  
 PT to isolate antimicrobial compounds, and in vaccines against S.

PT aureus infection

XX Claim 9; Page 940-941; 989pp; English.

CC The present sequence encodes a Staphylococcus aureus protein of unknown  
 CC function. The present sequence was isolated from a library of clones of  
 CC S. aureus WCUH 29 in Escherichia coli. The DNA sequence can be used in  
 CC the construction of ribozymes and antisense sequences to control the  
 CC expression of Staphylococcal genes. The DNA sequence is also useful  
 CC as a source of regulatory elements for the control of bacterial gene  
 CC expression. The encoded protein may be used to produce vaccines to  
 CC enable a host to produce specific antibodies with antibacterial action.  
 CC These vaccines and antibodies would protect a host against invasion by  
 CC S. aureus, and conditions relating to Staphylococcal infection, e.g.  
 CC Staphylococcal food poisoning, scaled skin syndrome, and toxic shock  
 CC syndrome.

XX SQ Sequence 1000 BP; 289 A; 235 C; 151 G; 313 T; 12 other;

alignment\_scores:  
 Quality: 9.00 Length: 9  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-701-711-2 x AAT84180/rev ..

Align seg 1/1 to reverse of: AAT84180 from: 1 to: 1000

717 ValArgProValIlePheIleGluGly 725  
 |||||  
 311 GTGCGACCCGTCATTTTATTGAAGGA 285

seq\_name: /SIDSL/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:AAV53520

seq\_documentation\_block:

ID AAV53520 standard; DNA; 1000 BP.

XX AC AAV53520;

XX DT 30-OCT-1998 (first entry)

XX DE DNA encoding a Exou protein.

XX Staphylococcus aureus protein; immune response induction; eye infection;  
 KW antibody production; i-cell immune response; gastrointestinal infection;  
 KW respiratory infection; inhibitor; bacterial infection; cardiac infection;  
 KW central nervous system; kidney infection; urinary tract infection;  
 KW antimicrobial compound identification; broad spectrum antibiotic;  
 KW therapy; ss.

XX OS Staphylococcus aureus.

XX PN EP841394-A2.

XX PD 13-MAY-1998.

XX PF 24-SEP-1997; 97EP-0307485.

XX PR 24-SEP-1996; 96US-0027032.

XX (SMIK ) SMITHKLINE BEECHAM CORP.  
 XX (SMIK ) SMITHKLINE BEECHAM PLC.

XX Black MT, Burnham MKR, Hodgson JE, Knowles DJC;  
 PI Lonetto MA, Nicholas RO, Pratt JM, Reichard RW, Rosenberg M;  
 PI Ward JM;

XX WPI; 1998-252940/23.

XX P-PSDB; AAW77731.

XX New nucleic acid sequences from Staphylococcus aureus WCUH29 -

PT useful in vaccines and for treatment of bacterial infections of e.g.  
PT respiratory tract and central nervous system  
XX  
PS Claim 1; Page 185; 390pp; English.  
XX  
CC This sequence encodes a *Staphylococcus aureus* protein that (based on  
CC homology with a *Rhizobium meliloti* protein) is a  
CC Exon protein, and represents a DNA sequence of the invention.  
CC The DNA sequences were isolated from *Staphylococcus aureus* WCHU29  
CC (NCIMB 40771). Host cells containing the DNA sequences are used to  
CC produce polypeptides or fragments. The proteins are used in the treatment  
CC of disease, for inducing an immune response by administering them, to  
CC produce antibody and/or T-cell immune response. Antagonists of the  
CC proteins are used for the inhibition of bacterial polypeptides.  
CC Conditions which may be treated include bacterial infections, especially  
CC respiratory, cardiac, gastrointestinal, central nervous, eye, kidney,  
CC urinary tract, skin, bones and joints. The proteins can also be used to  
CC identify antimicrobial compounds which are broad spectrum antibiotics,  
CC especially useful in the treatment of *H. pylori* infection.  
XX  
SQ Sequence 1000 BP; 289 A; 235 C; 151 G; 313 T; 12 other;

alignment\_scores:  
Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-701-711-2 x AAV53520/rev ..

Align seg 1/1 to reverse of: AAV53520 from: 1 to: 1000

717 ValArgProValIlePheIleGluGly 725  
|||||  
311 GTGCGACCGTCATTTTATTGAAGGA 285

seq\_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.AAF61065

seq\_documentation\_block:  
ID AAF61065 standard; DNA; 2019 BP.  
XX  
AC AAF61065;  
XX  
DT 16-MAY-2001 (first entry)  
XX  
DE P. putida KT2440-associated DNA ORF08500.  
XX  
KW Transgenic plant; detection; probe; amplification; vaccine carrier;  
KW microbial production strain; biological remediation; ds.  
XX  
OS *Pseudomonas putida*.  
XX  
PN DE19935088-A1.  
XX  
PD 01-FEB-2001.  
XX  
PF 27-JUL-1999; 99DE-1035088.  
XX  
PR 27-JUL-1999; 99DE-1035088.  
XX  
PA (TIGR-) TIGR INST GENOMIC RES.  
PA (QUIA-) QUIAGEN GMBH.  
PA (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.  
PA (DRFZ-) DRFZ DEUT KREBSFORSCHUNGSZENTRUM.  
PA (MEDI-) MEDIZINISCHE HOCHSCHULE HANNOVER.  
XX  
DR WPI; 2001-192469/20.  
XX  
PT New DNA sequences specific for *Pseudomonas putida* KT2440, useful as  
PT safe genetic engineering host, allow detection in presence of other  
PT related bacteria -  
XX

PS Claim 1a; Page 117-118; 158pp; German.

XX  
CC This invention describes novel DNA sequences (I) for specific detection  
CC of *Pseudomonas putida* KT2440. The invention also describes (1)  
CC recombinant expression vector containing (I); (2) prokaryotic or  
CC eukaryotic cells transformed or transfected with (I) or the vector of  
CC (1); (3) production of expression products by culturing cells of (2);  
CC (4) expression products, or their fragments, of (I) and synthetic  
CC proteins or peptides with the same sequences (A); (5) poly- or  
CC mono-clonal antibodies (Ab) that react specifically with (A); (6)  
CC hybridoma cells that produce the monoclonal Ab of (5); (7) transgenic  
CC plants that contain transformed or transfected cells of (2); (8)  
CC detecting KT2440 using a labeled (I) or Ab as probe; and (9) DNA chips  
CC carrying one or more (I). (I), and their fragments, are used as probes  
CC to detect and isolate full-length cDNAs and/or to amplify such cDNAs by  
CC polymerase chain reaction, and for production of transgenic plants. (I),  
CC or antibodies that recognize their expression products, are used for  
CC detecting the presence of KT2440, particularly in presence of other,  
CC even closely related, bacteria. KT2440 is one of the bacteria classified  
CC as safe, by the National Institutes of Health, for genetic engineering  
CC work, e.g. as microbial production strains, for biological remediation  
CC and as vaccine carriers. (I) are exclusive to KT2440 with no significant  
CC homology with sequences in other bacteria (specifically the closely  
CC related pathogen *P. aeruginosa*). Compared with other 'safe' bacteria, it  
CC has greater catabolic activity and better survival in, and adaptation to,  
CC the rhizosphere and soil.  
XX  
SQ Sequence 2019 BP; 404 A; 671 C; 603 G; 338 T; 3 other;

alignment\_scores:  
Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-701-711-2 x AAF61065 ..

Align seg 1/1 to: AAF61065 from: 1 to: 2019

368 AspGluValLeuArgArgGluMetArg 376  
|||||  
250 GACGAAGTGTCTGCTGCGAAATGCGC 276

seq\_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.AAL29004

seq\_documentation\_block:

ID AAL29004 standard; DNA; 50 BP.

XX  
AC AAL29004;

XX  
DT 24-JAN-2002 (first entry)

XX  
DE Human SNP oligonucleotide #2212.

XX  
KW Immunosuppressive; immunostimulatory; antiinflammatory; cytostatic;  
KW neuroprotective; antimicrobial; gene therapy; vaccine; amylase; cancer;  
KW amyloid protein; angiotensin; apoptosis related protein; cadherin;  
KW cyclin; polymerase; oncogene; histone; kinase; colony stimulating factor;  
KW complement related protein; cytochrome; kinesin; cytokine; interferon;  
KW interleukin; G-protein coupled receptor; thioesterase; inflammation;  
KW multifactorial disease; autoimmune disease; infection;  
KW nervous system disease; ss.

OS Homo sapiens.

XX WO200147944-A2.

PN 05-JUL-2001.

PD 28-DEC-2000; 2000WO-US35498.

XX 28-DEC-1999; 99US-0173419.

```
PR 27-DEC-2000; 2000US-0173419.
XX (CURA-) CURAGEN CORP.
PA
XX Shimkets RA, Leach M;
XX WPI; 2001-465210/50.
XX
XX Polymorphic nucleic acids encoding e.g. amylases, cyclins, polymerases,
PT oncogenes and histones, useful for diagnosing and treating, e.g.
PT cancer, autoimmune diseases and infections -
XX
XX Claim 1; Page 2015; 4143pp; English.
XX
XX The present invention relates to oligonucleotides encoding polymorphic
CC variants of proteins related to amylases, amyloid proteins, angiotensin,
CC apoptosis related proteins, cadherin, cyclin, polymerase, oncogenes,
CC histones, kinases, colony stimulating factors, complement related
CC proteins, cytochromes, kinesins, cytokines, interferons, interleukins,
CC G-protein coupled receptors and thioesterases. The present sequence is
CC one such oligonucleotide. The oligonucleotides and the peptides encoded
CC by them may be used in the prevention, diagnosis and treatment of
CC diseases associated with inappropriate expression of the proteins listed
CC above. Disorders that may be prevented, diagnosed and/or treated include
CC multifactorial diseases with a genetic component, such as autoimmune
CC diseases (e.g. rheumatoid arthritis, multiple sclerosis, diabetes,
CC systemic lupus erythematosus and Grave's disease), inflammation, cancer
CC (e.g. cancers of the bladder, brain, breast, colon and kidney,
CC leukaemia), diseases of the nervous system and an infection of pathogenic
CC organisms.
XX
XX Sequence 50 BP; 8 A; 11 C; 11 G; 20 T; 0 other;

alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AAL29004 ..
Align seg 1/1 to: AAL29004 from: 1 to: 50

488 GlyValSerGlnSerLeuSerGly 495
|||||
22 GGAGTTTCTCAGAGCCTCTGTT 45

seq_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT: AAD07056

seq_documentation_block:
ID AAD07056 standard; DNA; 310 BP.
XX
AC AAD07056;
XX
DT 06-AUG-2001 (first entry)
XX
DE Human DNA #1 related to the invention.
XX
XX Human; gene therapy; angiogenesis; vascularisation; cardiac hypertrophy;
KW congestive heart failure; CHF; myocardial infarction; aortic stenosis;
KW valvular regurgitation; vascular trauma; wound; burn; cancer; vaccine;
KW endothelial disease; ds.
XX
OS Homo sapiens.
XX
PN WO200125433-A2.
XX
PD 12-APR-2001.
XX
PF 05-OCT-2000; 2000WO-US27512.
XX
PR 07-OCT-1999; 99US-0158587.
```

```
PR 28-OCT-1999; 99US-0162611.
XX (GETH ) GENENTECH INC.
PA (CURA-) CURAGEN CORP.
XX
XX Gerritsen ME, Goddard A, Grimaldi JC, Mehraban F;
XX WPI; 2001-367229/38.
XX
XX Polypeptides critical for angiogenesis and vascularization, and the
PT nucleic acids encoding them, useful for treating conditions related to
PT inappropriate vascularization and angiogenesis -
XX
XX Disclosure; Page 187; 189pp; English.
XX
XX The patent discloses novel polypeptides critical for angiogenesis and
CC vascularisation and nucleic acids encoding them. The sequences of the
CC present invention are used in the prevention, treatment and diagnosis
CC of diseases associated with inappropriate angiogenesis and/or neo- or
CC cardio-vascularisation. They are used in the diagnosis, prevention
CC or treatment of angiogenic disorders such as cardiac hypertrophy,
CC congestive heart failure (CHF), myocardial infarction, aortic stenosis,
CC valvular regurgitation and vascular trauma such as wounds, burns or
CC surgery. They are also used in the treatment of cancer and endothelial
CC diseases. The proteins of the invention are used as antigens in the
CC production of antibodies against the angiogenesis and vascularisation
CC proteins and in assays to identify modulators (agonists and antagonists)
CC of their expression and activity. The antibodies and antagonists are
CC used to down regulate expression and activity of the protein of the
CC invention. The antibodies are also used as diagnostic agents for
CC detecting the presence of the proteins of the invention in samples.
CC The sequences of the invention are also used as vaccines and in gene
CC therapy. The present sequence is human DNA related to the invention.
XX
XX Sequence 310 BP; 83 A; 83 C; 63 G; 80 T; 1 other;

alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AAD07056 ..
Align seg 1/1 to: AAD07056 from: 1 to: 310

488 GlyValSerGlnSerLeuSerGly 495
|||||
158 GGAGTATCTCAAAGCCTTTCAGGC 181

seq_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT: AAK60152

seq_documentation_block:
ID AAK60152 standard; cDNA; 388 BP.
XX
AC AAK60152;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:5212.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
XX 09-AUG-2001.
XX
PD 17-JAN-2001; 2001WO-US01354.
XX
PR
```

PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236602.

PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI: 2001-483426/52.

P-PSDB; AAM87371.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
useful for preventing, diagnosing and/or treating cancers and



```
PT metastasis -
XX Claim 1; SEQ ID NO 5212; 3071pp + Sequence Listing; English.
PS
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 388 BP; 72 A; 94 C; 119 G; 100 T; 3 other;

alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AAK60152/rev ..
Align seg 1/1 to reverse of: AAK60152 from: 1 to: 388

115 GluGlyLeuGlnGluGlyLeuLys 122
|||||
361 GAAGGGTTACAGAGAGGTTTAAAG 338

seq_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AAK56055

seq_documentation_block:
ID AAK56055 standard; DNA; 413 BP.
XX
AC AAK56055;
XX
DT 25-JAN-2001 (first entry)
XX
DE Eucalyptus grandis transcription factor DNA sequence #186.
XX
KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;
KW type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
OS Eucalyptus grandis.
XX
PN W0200053724-A2.
XX
PD 14-SEP-2000.
XX
PF 09-MAR-2000; 2000WO-US06112.
XX
PR 11-MAR-1999; 99US-0266513.
PR 18-AUG-1999; 99US-0149485.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Wood M, McGrath A, Shenk MA, Glenn M;
XX WPI; 2000-579369/54.
XX

New isolated polynucleotide encoding a plant transcription factor for
producing a plant e.g. a woody plant, preferably eucalyptus or pine,
having modified gene expression or modified activity of a polypeptide
-
Claim 1; Page 93; 747pp; English.
XX
CC The present invention relates to novel plant transcription factors from
CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC sequence for one such transcription factor. The transcription factor may
CC be used to produce a plant having modified gene expression such as a
CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC mahogany species or to modify the activity of a polypeptide in a plant.
CC The transcription factors of the present invention are members from the
CC following families of regulatory proteins: bZIP, bZIP family of G-box
CC binding factors, basic helix-loop-helix zipper,
CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
XX
SQ Sequence 413 BP; 99 A; 92 C; 120 G; 102 T; 0 other;

alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AAC56055/rev ..
Align seg 1/1 to reverse of: AAC56055 from: 1 to: 413

429 SerGlySerSerThrIleAlaAla 436
|||||
311 TCAGGTAGTTCAACCATCGCAGCC 288

seq_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AAK03891

seq_documentation_block:
ID AAC03891 standard; cDNA; 447 BP.
XX
AC AAC03891;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 3889.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST ) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
XX
PR P-PSDB; AAG03885.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
```

PS Claim 1; SEQ ID 3889; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from

CC mRNAs encoding secreted proteins. An ORF has been identified within the

CC sequence. The 5' ESTs were prepared from total human RNAs or polyA<sup>+</sup> RNAs

CC derived from 30 different tissues. EST sequences usually correspond

CC mainly to the 3' untranslated region (UTR) of the mRNA because they are

CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not

CC well suited for isolating cDNA sequences derived from the 5' ends of

CC mRNAs and even in those cases where longer cDNA sequences have been

CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from

CC mRNAs with intact 5' ends and can therefore be used to obtain full length

CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,

CC gene therapy and chromosome mapping procedures. They are used to obtain

CC upstream regulatory sequences and to design expression and secretion

XX vectors.

SQ Sequence 447 BP; 111 A; 130 C; 119 G; 84 T; 3 other;

alignment\_scores:

Quality: 8.00 Length: 8

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-701-711-2 x AAC03891 ..

Align seg 1/1 to: AAC03891 from: 1 to: 447

113 ProlylsGluGlyLeuGlnGluGly 120

|||||

78 CCCAAGAAGGGCTCCCAAGAGGC 101

seq\_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:AAQ79427

seq\_documentation\_block:

ID\_ AAQ79427 standard; cDNA; 450 BP.

XX AC AAQ79427;

XX 03-JUN-1995 (first entry)

XX Human islet inducible nitric oxide synthase (iNOS) cDNA 3' end.

XX Nitric oxide synthase; diabetes mellitus; islet; ss.

XX Homo sapiens.

XX WO9424269-A.

XX 27-OCT-1994.

XX 11-APR-1994; 94WO-DK00146.

XX 16-APR-1993; 93DK-0000433.

XX (NOVO ) NOVO-NORDISK AS.

XX Karlsen AE;

XX WPI; 1994-341851/42.

XX New DNA encoding inducible pancreatic islet nitric oxide synthase

PT - and related vectors and transformed cells, useful for

PT identifying specific inhibitors for treatment or prevention of

PT insulin dependent diabetes mellitus

XX Disclosure; Page 26-27; 36pp; English.

XX Human islets were incubated in a mixture of IL-1, TNF-alpha

CC and IFN-gamma. The human islet iNOS was cloned by RT-PCR on

CC the isolated mRNA with primers based on the human

CC hepatocyte sequence. The sequence of the primers is given in

CC AAQ79424 and AAQ79425. The clone human islet iNOS was sequenced.

CC The 5' UTR and translated human islet iNOS is given in AAQ79426,

CC and the 3' end in AAQ79427.

SQ Sequence 450 BP; 89 A; 142 C; 116 G; 103 T; 0 other;

alignment\_scores:

Quality: 8.00 Length: 8

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-701-711-2 x AAQ79427/rev ..

Align seg 1/1 to reverse of: AAQ79427 from: 1 to: 450

115 GluGlyLeuGlnGluGlyLeuLys 122

|||||

146 GAGGCGCTACAGGAGGGCTTAAG 123

seq\_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:ABA62517

seq\_documentation\_block:

ID\_ ABA62517 standard; DNA; 511 BP.

XX AC ABA62517;

XX 01-FEB-2002 (first entry)

XX Human foetal liver single exon nucleic acid probe #10822.

XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

XX Homo sapiens.

XX WO200157277-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00669.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human fetal liver -

XX Claim 1; SEQ ID NO 10822; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for

CC measuring human gene expression in a sample derived from human foetal

CC liver. The single exon nucleic acid probes may be used for predicting,

CC measuring and displaying gene expression in samples derived from human

CC fetal liver. The present sequence is a single exon nucleic acid

CC probe of the invention.

CC Note: The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly

CC from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).

XX Sequence 511 BP; 114 A; 129 C; 121 G; 147 T; 0 other;

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-701-711-2 x ABA62517/rev ..  
Align seg 1/1 to reverse of: ABA62517 from: 1 to: 511

47 GluSerLeuGlnSerValLeuPro 54  
|||||  
462 GAGAGCCTCCAGTCTGTCCTTCT 439

seq\_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:ABA29847

seq\_documentation\_block:  
ID ABA29847 standard; DNA; 511 BP.  
XX  
AC ABA29847;  
XX  
DT 23-JAN-2002 (first entry)  
XX  
DE Probe #8313 for gene expression analysis in human heart cell sample.  
XX  
KW Human; gene expression; heart; microarray; vascular system; probe;  
KW cardiovascular disease; hypertension; cardiac arrhythmia;  
KW congenital heart disease; ss.

XX OS Homo sapiens.  
XX  
XX PN WO200157274-A2.  
XX  
XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00666.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-488899/53.

XX PT Single exon nucleic acid probes for analyzing gene expression in human hearts -

XX PS Claim 1; SEQ ID No 8313; 530pp; English.

XX CC The present invention relates to single exon nucleic acid probes for measuring human gene expression in a sample derived from human heart. The present sequence is one such probe. The probes may be used for predicting, measuring and displaying gene expression in samples derived from the human heart via microarrays. By measuring gene expression, the probes are useful for predicting, diagnosing, grading, staging, monitoring and prognosing diseases of the human heart and vascular system e.g. cardiovascular disease, hypertension, cardiac arrhythmias and congenital heart disease.

XX CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX XX Sequence 511 BP; 114 A; 129 C; 121 G; 147 T; 0 other;

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-701-711-2 x ABA29847/rev ..  
Align seg 1/1 to reverse of: ABA29847 from: 1 to: 511

47 GluSerLeuGlnSerValLeuPro 54  
|||||  
462 GAGAGCCTCCAGTCTGTCCTTCT 439

seq\_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AAK10872

seq\_documentation\_block:  
ID AAK10872 standard; DNA; 511 BP.

XX AC AAK10872;

XX DT 05-NOV-2001 (first entry)

XX DE Human brain expressed single exon probe SEQ ID NO: 10863.

XX KW Human; brain expressed exon; gene expression analysis; probe;

XX KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;

XX KW epilepsy; cancer; ss.

XX OS Homo sapiens.

XX PN WO200157275-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00667.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-483446/52.

XX PT Single exon nucleic acid probes for analyzing gene expression in human brains -

XX PS Example 4; SEQ ID NO: 10863; 650pp + Sequence Listing; English.

XX CC The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human brain. They can be used to measure gene expression in brain cell samples, which may enable the diagnosis and improved treatment of nervous system diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, epilepsy and cancers. The present sequence is one of the probes of the invention.

XX SQ Sequence 511 BP; 114 A; 129 C; 121 G; 147 T; 0 other;

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

```
alignment_block:
US-09-701-711-2 x AAK10872/rev ..
Align seg 1/1 to reverse of: AAK10872 from: 1 to: 511
47 GluSerLeuGlnSerValleuPro 54
|||||
462 GAGAGCCTCCAGTCGTCTCCT 439

seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AAK36734
seq_documentation_block:
ID AAK36734 standard; DNA; 511 BP.
XX
AC AAK36734;
XX
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 11291.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
FN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow.
XX
PS Example 4; SEQ ID NO: 11291; 658pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
SQ Sequence 511 BP; 114 A; 129 C; 121 G; 147 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AAK36734/rev ..
Align seg 1/1 to reverse of: AAK36734 from: 1 to: 511
47 GluSerLeuGlnSerValleuPro 54
|||||
462 GAGAGCCTCCAGTCGTCTCCT 439
```

```
seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA117584
seq_documentation_block:
ID AA117584 standard; DNA; 511 BP.
XX
AC AA117584;
XX
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #7517 for gene expression analysis in human cervical cell sample.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
OS Homo sapiens.
XX
FN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00670.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells.
XX
PS Claim 25; SEQ ID No 7517; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENPs). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 511 BP; 114 A; 129 C; 121 G; 147 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AA117584/rev ..
Align seg 1/1 to reverse of: AA117584 from: 1 to: 511
47 GluSerLeuGlnSerValleuPro 54
|||||
462 GAGAGCCTCCAGTCGTCTCCT 439

seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA142500
seq_documentation_block:
```

```
ID AAI42500 standard; DNA; 511 BP.
XX
AC AAI42500;
XX
DT 17-OCT-2001. (first entry)
XX
DE Probe #1186 used to measure gene expression in human placenta sample.
XX
KW Probe: microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 11186; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 511 BP; 114 A; 129 C; 121 G; 147 T; 0 other;

alignment_scores:
    Quality: 8.00      Length: 8
    Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AAI42500/rev ..

Align seg 1/1 to reverse of: AAI42500 from: 1 to: 511

47 GluSerLeuGlnSerValLeuPro 54
|||||
462 GAGAGCCTCCAGTCTGCTCCTCT 439

seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AAS42998

seq_documentation_block:
ID AAS42998 standard; cDNA; 583 BP.
XX
AC AAS42998;
XX
DT 18-DEC-2001 (first entry)
XX
DE DNA encoding G protein-coupled receptor, nGPCR-2102.
XX
KW Human; mental disorder; thyroid disease; renal failure; anorexia;
```

```
KW inflammatory condition; Crohn's disease; rheumatoid arthritis; HIV;
KW autoimmune disorder; schizophrenia; migraine; stroke; dementia; obesity;
KW depression; Parkinson's disease; Alzheimer's disease; viral infection;
KW Huntington's disease; human immunodeficiency virus; type 2 diabetes;
KW anorexia; hypotension; hypertension; thrombosis; myocardial infarction;
KW atherosclerosis; cancer; sexual dysfunction; G protein-coupled receptor;
KW nGPCR; ss.
XX
OS Homo sapiens.
XX
PN WO200162924-A2.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-US05989.
XX
PR 24-FEB-2000; 2000US-0184602.
PR 24-FEB-2000; 2000US-0184604.
PR 24-FEB-2000; 2000US-0184606.
PR 24-FEB-2000; 2000US-0184689.
PR 24-FEB-2000; 2000US-0184690.
PR 24-FEB-2000; 2000US-0184710.
PR 24-FEB-2000; 2000US-0184712.
PR 24-FEB-2000; 2000US-0184715.
PR 24-FEB-2000; 2000US-0184716.
PR 24-FEB-2000; 2000US-0184725.
PR 24-FEB-2000; 2000US-0184822.
XX
PA (PHAA ) PHARMACIA & UPJOHN CO.
XX
PI Vogeli G, Wood LS, Parodi LA, Lind P;
XX
DR WPI; 2001-570632/64.
XX
DR P-PSDB; AAU25688.
XX
PT Novel nucleic acid and encoded nGPCR-x, used to screen for compounds
PT for use in the treatment of mental disorders, such as Alzheimer's
PT disease, or Parkinson's disease -
XX
PS Claim 3; Page 81; 263pp; English.
XX
CC The invention relates to novel isolated human G protein-coupled
CC receptors (nGPCR-x). The nGPCR-x can be used for screening compounds
CC which can be used to treat mental disorders, thyroid disease, renal
CC failure, inflammatory conditions such as Crohn's disease, rheumatoid
CC arthritis, autoimmune disorders, schizophrenia, migraine, stroke,
CC dementia, depression, Parkinson's disease, Alzheimer's disease, and
CC Huntington's disease. They may also be used for treating viral infections
CC such as human immunodeficiency virus (HIV), type 2 diabetes, obesity,
CC anorexia, hypotension, hypertension, thrombosis, myocardial infarction,
CC atherosclerosis, cancer, and sexual dysfunction. AAS42927-AAS43036
CC represent the coding sequences of novel human G protein-coupled
CC receptors, nGPCR-2031 to nGPCR-2140 respectively, as described in the
CC invention.
XX
SQ Sequence 583 BP; 190 A; 86 C; 102 G; 205 T; 0 other;
```

```
alignment_scores:
    Quality: 8.00      Length: 8
    Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

```
alignment_block:
US-09-701-711-2 x AAS42998 ..

Align seg 1/1 to: AAS42998 from: 1 to: 583
```

```
310 ThrThrAsnAsnIleSerThrIys 317
|||||
22 ACAACAAACAATATTTCTACTAAA 45
```

```
seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:AAV35370
```

```
seq_documentation_block:
ID  AAV35370 standard; DNA; 604 BP.
XX
AC  AAV35370;
XX
DT  09-OCT-1998 (first entry)
XX
DE  Human INOS genomic DNA fragment #1.
XX
KW  Inducible nitric oxide synthase; hINOS; human; regulator; expression;
KW  treatment; inflammation; sepsis; antitumour agent; inhibitor; ds.
XX
OS  Homo sapiens.
XX
PN  WO9812313-A1.
XX
PD  26-MAR-1998.
XX
PF  18-SEP-1997; 97WO-JP03303.
XX
PR  20-SEP-1996; 96JP-0250697.
XX
PA  (SUNR ) SUNTORY LTD.
XX
PI  Nunokawa Y, Oikawa S, Tanaka S;
XX
PS  WPI; 1998-230314/20.
XX
PT  Screening for regulators of human inducible nitric oxide synthase
PT  expression - using human cell line transformed with a reporter gene
PT  flanked by 5'-promoter and 3'-non-translated regions of the hINOS
PT  gene
XX
PS  Disclosure; Fig 2; 56pp; Japanese.
XX
CC  This sequence is a fragment of human inducible nitric oxide synthase
CC  (hINOS) which can be used to screen for potential hINOS expression
CC  regulators. Such regulators have an effect on the modification of the
CC  expression (in the presence of a cytokine inducer) of a reporter gene in
CC  a human cell line transformed with an expression vector containing (in
CC  order from 5'-end): the 5'-flanking region of the hINOS gene containing
CC  544 bases, the reporter gene and the 3'-untranslated region of the hINOS
CC  gene. Suitable reporter genes include chloramphenicol acetyltransferase
CC  (CAT), beta-galactosidase (beta -gal) and luciferase. Compounds which
CC  influence the expression of the hINOS gene are useful for the treatment
CC  of inflammation and sepsis, as antitumour agents and for the inhibition
CC  of re-angio-stenosis, and other conditions involving abnormalities of
CC  hINOS expression.
XX
SQ  Sequence 604 BP; 158 A; 150 C; 146 G; 150 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AAV35370 ..
Align seg 1/1 to: AAV35370 from: 1 to: 604
115 GluGlyLeuGlnGluGlyLeuLys 122
|||||
107 GAGGGCCTACAGGAGGGGTTAAAG 130

seq_name: /SIDS1/gcdata/geneseq/geneseq-embl/NA1999.DAT:AAZ32098
seq_documentation_block:
ID  AAZ32098 standard; DNA; 604 BP.
XX
AC  AAZ32098;
XX
DT  13-MAR-2001 (first entry)
```

```
XX 11-JAN-2000 (first entry)
DT
DE HINOS exon 26 nucleotide sequence.
XX
KW Human induced nitrogen monoxide synthase; hINOS; untranslated region;
KW NF-kappaB; expression; activation; regulation; ds.
XX
OS Homo sapiens.
XX
PN JP11266872-A.
XX
PD 05-OCT-1999.
XX
PF 20-MAR-1998; 98JP-0090664.
XX
PR 20-MAR-1998; 98JP-0090664.
XX
PA (SUNR ) SUNTORY LTD.
XX
PI WPI; 1999-613779/53.
XX
PT Method of screening a substance controlling activation of NF-kappaB -
PT having increased sensitivity
XX
PS Disclosure; Fig 2; 16pp; Japanese.
XX
CC The present invention describes an expression-controlling sequence
CC containing a NF-kappaB recognising sequence, at least part of
CC 3'-untranslated region (3'-UTR) of human-induced nitrogen monoxide
CC synthase (hINOS) and at least part of 3'-flanking region. Also described
CC are: (a) an expression vector containing the above expression-controlling
CC sequence; (b) an expression vector containing (1) 5'-flanking region of
CC hINOS gene containing a promoter region, (2) a reporter gene and (3) the
CC above expression-controlling sequence in this order from the 5'-side;
CC (c) a cell transformed by the above expression vector; (d) a method for
CC screening a substance controlling activation of NF-kappaB in which a cell
CC which has the above expression-controlling sequence and can detect the
CC activation of NF-kappaB is treated with a sample to observe the change in
CC the expressed amount of the reporter gene; (e) a kit for screening a
CC substance controlling activation of NF-kappaB containing the above cell;
CC (f) a compound controlling activation of NF-kappaB prepared by using the
CC above screening method; and (g) a drug composition for treating diseases
CC caused by activation of NF-kappaB containing, with the above compound as
CC the active component. The method can evaluate a compound controlling
CC activation of NF-kappaB easily in a high sensitivity. The present
CC sequence is used in the exemplification of the present invention.
XX
SQ Sequence 604 BP; 158 A; 150 C; 146 G; 150 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AAZ32098 ..
Align seg 1/1 to: AAZ32098 from: 1 to: 604
115 GluGlyLeuGlnGluGlyLeuLys 122
|||||
107 GAGGGCCTACAGGAGGGGTTAAAG 130

seq_name: /SIDS1/gcdata/geneseq/geneseq-embl/NA2000.DAT:AAFI6109
seq_documentation_block:
ID  AAFI6109 standard; cDNA; 842 BP.
XX
AC  AAFI6109;
XX
DT  13-MAR-2001 (first entry)
```

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:544.  
DE  
XX  
XX Human; prostate cancer; prostate cancer antigen; detection; diagnosis;  
KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;  
KW vulnary; gastrointestinal; nephrotropic; antineoplastic; gynaecological;  
KW antibacterial; gene therapy; neural; immune; reproductive; renal;  
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;  
KW wound; infectious disease; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
XX W0200055174-A1.  
PN  
XX  
XX 21-SEP-2000.  
PD  
XX  
XX 08-MAR-2000; 2000WO-US05988.  
PF  
XX  
XX 12-MAR-1999; 99US-0124270.  
PR  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
XX  
XX (ROSE/) ROSEN C A.  
PI  
XX  
XX Rosen CA, Ruben SM;  
XX  
XX WPI; 2000-587513/55.  
DR  
XX  
XX P-PSDB; AAB56906.  
DR  
XX  
XX Prostate cancer associated gene sequences, referred to as prostate  
PT cancer antigens, useful for treatment, prevention, and diagnosis of  
PT disorders such as prostate cancer -  
XX  
XX  
PS Claim 1; Page 1004; 2338pp; English.  
XX  
XX AAF15566 to AAF16505 encode the human prostate cancer associated  
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.  
CC The prostate cancer antigens can have neuroprotective, cytostatic,  
CC cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,  
CC nephrotropic, antineoplastic, gynaecological and antibacterial activities,  
CC and can be used in gene therapy. The prostate cancer antigen  
CC polynucleotides may be used for detection of prostate cancer  
CC identification, as chromosome markers, and for numerous other, diagnostic  
CC or research purposes. The prostate cancer antigens may be used to treat  
CC disorders such as neural, immune, muscular, reproductive,  
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative  
CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to  
CC AAB57303 represent sequences used in the exemplification of the present  
CC invention.  
XX  
XX  
SQ Sequence 842 BP; 257 A; 201 C; 205 G; 179 T; 0 other;

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-701-711-2 x AAF16109 ..  
Align seg 1/1 to: AAF16109 from: 1 to: 842

113 ProLysGluGlyLeuGlnGluGly 120  
|||||  
101 CCCAAGAGGGCTCCAGAGGC 124

seq\_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT: AAS81839

seq\_documentation\_block:

ID AAS81839 standard; cDNA; 863 BP.

XX AAS81839;

AC

XX

DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #17643.  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
XX W0200175067-A2.  
PN  
XX  
XX 11-OCT-2001.  
PD  
XX  
XX 30-MAR-2001; 2001WO-US08631.  
PF  
XX  
XX 31-MAR-2000; 2000US-0540217.  
PR  
XX  
XX 23-AUG-2000; 2000US-0649167.  
PR  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX  
XX Dmanac RT, Liu C, Tang YT;  
XX  
XX WPI; 2001-639362/73.  
DR  
XX  
XX P-PSDB; ABG17652.  
DR  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
XX  
PS Claim 1; SEQ ID No 17643; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_sequences.  
XX  
SQ Sequence 863 BP; 222 A; 214 C; 231 G; 196 T; 0 other;

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-701-711-2 x AAS81839 ..  
Align seg 1/1 to: AAS81839 from: 1 to: 863

370 ValLeuArgArgGluMetArgGln 377  
|||||  
6 GTCCCTGCGTCCGAATGCGTCAG 29

seq\_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT: AAV35383

seq\_documentation\_block:  
ID AAV35383 standard; DNA; 1026 BP.  
AC AAV35383;  
XX  
DT 09-OCT-1998 (first entry)  
XX  
DE Human iNOS genomic DNA fragment 3'-end #1.  
XX  
KW Inducible nitric oxide synthase; hINOS; human; regulator; expression;  
KW treatment; inflammation; sepsis; antitumour agent; inhibitor; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO9812313-A1.  
XX  
PD 26-MAR-1998.  
XX  
PF 18-SEP-1997; 97WO-JP03303.  
XX  
PR 20-SEP-1996; 96JP-0250697.  
XX  
PA (SUNR ) SUNTORY LTD.  
XX  
PI Nunokawa Y, Oikawa S, Tanaka S;  
XX  
DR WPI; 1998-230314/20.  
XX  
XX Screening for regulators of human inducible nitric oxide synthase  
PT expression - using human cell line transformed with a reporter gene  
PT flanked by 5'-promoter and 3'-non-translated regions of the hINOS  
PT gene  
XX  
PS Disclosure; Page 28-29; 56pp; Japanese.  
XX  
XX This sequence is a 3'-end fragment of human inducible nitric oxide  
CC synthase (hINOS) which can be used to screen for potential hINOS  
CC expression regulators. Such regulators have an effect on the modification  
CC of the expression (in the presence of a cytokine inducer) of a reporter  
CC gene in a human cell line transformed with an expression vector  
CC containing (in order from 5'-end): the 5'-flanking region of the hINOS  
CC gene containing 544 bases, the reporter gene and the 3'-untranslated  
CC region of the hINOS gene. Suitable reporter genes include chloramphenicol  
CC acetyltransferase (CAT), beta-galactosidase (beta -Gal) and luciferase.  
CC Compounds which influence the expression of the hINOS gene are useful for  
CC the treatment of inflammation and sepsis, as antitumour agents and for  
CC the inhibition of re-angio-stenosis, and other conditions involving  
CC abnormalities of hINOS expression.  
XX  
SQ Sequence 1026 BP; 219 A; 259 C; 299 G; 249 T; 0 other;

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-701-711-2 x AAV35383 ..  
Align seg 1/1 to: AAV35383 from: 1 to: 1026

115 GluGlyLeuGlnGluGlyLeuLys 122  
|||||  
2 GAGGGCCTACAGGAGGGGTTAAAG 25

seq\_name: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:AAZ32093  
seq\_documentation\_block:  
ID AAZ32093 standard; DNA; 1026 BP.  
XX  
AC AAZ32093;  
XX

DT 11-JAN-2000 (first entry)  
XX  
DE hINOS 3'-untranslated region SEQ ID NO:13.  
XX  
KW Human induced nitrogen monoxide synthase; hINOS; untranslated region;  
KW NF-kappaB; expression; activation; regulation; ds.  
XX  
OS Homo sapiens.  
XX  
PN JP11266872-A.  
XX  
PD 05-OCT-1999.  
XX  
PF 20-MAR-1998; 98JP-0090664.  
XX  
PR 20-MAR-1998; 98JP-0090664.  
XX  
PA (SUNR ) SUNTORY LTD.  
XX  
DR WPI; 1999-613779/53.  
XX  
PT Method of screening a substance controlling activation of NF-kappaB -  
PT having increased sensitivity  
XX  
PS Example 1; Page 10-11; 16pp; Japanese.  
XX  
XX The present invention describes an expression-controlling sequence  
CC containing a NF-kappaB recognising sequence, at least part of  
CC 3'-untranslated region (3'-UTR) of human-induced nitrogen monoxide  
CC synthase (hINOS) and at least part of 3'-flanking region. Also described  
CC are: (a) an expression vector containing the above expression-controlling  
CC sequence; (b) an expression vector containing (1) 5'-flanking region of  
CC hINOS gene containing a promoter region, (2) a reporter gene and (3) the  
CC above expression-controlling sequence in this order from the 5'-side;  
CC (c) a cell transformed by the above expression vector; (d) a method for  
CC screening a substance controlling activation of NF-kappaB in which a cell  
CC which has the above expression-controlling sequence and can detect the  
CC activation of NF-kappaB is treated with a sample to observe the change in  
CC the expressed amount of the reporter gene; (e) a kit for screening a  
CC substance controlling activation of NF-kappaB containing the above cell;  
CC (f) a compound controlling activation of NF-kappaB prepared by using the  
CC above screening method; and (g) a drug composition for treating diseases  
CC caused by activation of NF-kappaB containing, with the above compound as  
CC the active component. The method can evaluate a compound controlling  
CC activation of NF-kappaB easily in a high sensitivity. The present  
CC sequence represents hINOS 3'-untranslated region from the present  
XX invention.  
SQ Sequence 1026 BP; 219 A; 259 C; 299 G; 249 T; 0 other;

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-701-711-2 x AAZ32093 ..  
Align seg 1/1 to: AAZ32093 from: 1 to: 1026

115 GluGlyLeuGlnGluGlyLeuLys 122  
|||||  
2 GAGGGCCTACAGGAGGGGTTAAAG 25

seq\_name: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AAC38511  
seq\_documentation\_block:  
ID AAC38511 standard; DNA; 1143 BP.  
XX  
AC AAC38511;  
XX  
DT 17-OCT-2000 (first entry)



XX Arabidopsis thaliana DNA fragment SEQ ID NO: 21235.  
XX DE  
XX XX  
KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
XX metabolic pathway; promoter; termination sequence; ss.  
XX OS  
XX Arabidopsis thaliana.  
XX EP1033405-A2.  
XX PD  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-0301439.  
XX XX  
XX 25-FEB-1999; 99US-0121825.  
XX PR 05-MAR-1999; 99US-0123180.  
XX PR 09-MAR-1999; 99US-0123548.  
XX PR 23-MAR-1999; 99US-0125788.  
XX PR 29-MAR-1999; 99US-0126264.  
XX PR 01-APR-1999; 99US-0127462.  
XX PR 06-APR-1999; 99US-0128234.  
XX PR 08-APR-1999; 99US-0128714.  
XX PR 16-APR-1999; 99US-0129845.  
XX PR 19-APR-1999; 99US-0130077.  
XX PR 21-APR-1999; 99US-0130449.  
XX PR 23-APR-1999; 99US-0130510.  
XX PR 23-APR-1999; 99US-0130891.  
XX PR 28-APR-1999; 99US-0131449.  
XX PR 30-APR-1999; 99US-0132048.  
XX PR 30-APR-1999; 99US-0132407.  
XX PR 04-MAY-1999; 99US-0132484.  
XX PR 05-MAY-1999; 99US-0132485.  
XX PR 06-MAY-1999; 99US-0132486.  
XX PR 07-MAY-1999; 99US-0132487.  
XX PR 11-MAY-1999; 99US-0132863.  
XX PR 14-MAY-1999; 99US-0134256.  
XX PR 14-MAY-1999; 99US-0134218.  
XX PR 14-MAY-1999; 99US-0134219.  
XX PR 14-MAY-1999; 99US-0134221.  
XX PR 14-MAY-1999; 99US-0134370.  
XX PR 18-MAY-1999; 99US-0134768.  
XX PR 19-MAY-1999; 99US-0134941.  
XX PR 20-MAY-1999; 99US-0135124.  
XX PR 21-MAY-1999; 99US-0135353.  
XX PR 24-MAY-1999; 99US-0135623.  
XX PR 25-MAY-1999; 99US-0136021.  
XX PR 27-MAY-1999; 99US-0136392.  
XX PR 28-MAY-1999; 99US-0136782.  
XX PR 01-JUN-1999; 99US-0137222.  
XX PR 03-JUN-1999; 99US-0137528.  
XX PR 04-JUN-1999; 99US-0137502.  
XX PR 07-JUN-1999; 99US-0137724.  
XX PR 08-JUN-1999; 99US-0138094.  
XX PR 10-JUN-1999; 99US-0138540.  
XX PR 10-JUN-1999; 99US-0138847.  
XX PR 14-JUN-1999; 99US-0139119.  
XX PR 16-JUN-1999; 99US-0139452.  
XX PR 16-JUN-1999; 99US-0139453.  
XX PR 17-JUN-1999; 99US-0139492.  
XX PR 18-JUN-1999; 99US-0139454.  
XX PR 18-JUN-1999; 99US-0139455.  
XX PR 18-JUN-1999; 99US-0139456.  
XX PR 18-JUN-1999; 99US-0139457.  
XX PR 18-JUN-1999; 99US-0139458.  
XX PR 18-JUN-1999; 99US-0139459.  
XX PR 18-JUN-1999; 99US-0139460.  
XX PR 18-JUN-1999; 99US-0139461.  
XX PR 18-JUN-1999; 99US-0139462.  
XX PR 18-JUN-1999; 99US-0139463.  
XX PR 18-JUN-1999; 99US-0139750.  
XX PR 18-JUN-1999; 99US-0139763.  
XX PR 21-JUN-1999; 99US-0139817.  
XX PR 22-JUN-1999; 99US-0139899.  
XX PR 23-JUN-1999; 99US-0140353.  
XX PR 23-JUN-1999; 99US-0140354.  
XX PR 24-JUN-1999; 99US-0140695.  
XX PR 28-JUN-1999; 99US-0140823.  
XX PR 29-JUN-1999; 99US-0140991.  
XX PR 30-JUN-1999; 99US-0141287.  
XX PR 01-JUL-1999; 99US-0141842.  
XX PR 02-JUL-1999; 99US-0142154.  
XX PR 02-JUL-1999; 99US-0142055.  
XX PR 06-JUL-1999; 99US-0142390.  
XX PR 08-JUL-1999; 99US-0142803.  
XX PR 09-JUL-1999; 99US-0142920.  
XX PR 12-JUL-1999; 99US-0142977.  
XX PR 13-JUL-1999; 99US-0143542.  
XX PR 14-JUL-1999; 99US-0143624.  
XX PR 15-JUL-1999; 99US-0144005.  
XX PR 16-JUL-1999; 99US-0144085.  
XX PR 16-JUL-1999; 99US-0144086.  
XX PR 19-JUL-1999; 99US-0144325.  
XX PR 19-JUL-1999; 99US-0144331.  
XX PR 19-JUL-1999; 99US-0144332.  
XX PR 19-JUL-1999; 99US-0144333.  
XX PR 19-JUL-1999; 99US-0144334.  
XX PR 19-JUL-1999; 99US-0144335.  
XX PR 20-JUL-1999; 99US-0144352.  
XX PR 20-JUL-1999; 99US-0144632.  
XX PR 20-JUL-1999; 99US-0144884.  
XX PR 21-JUL-1999; 99US-0144814.  
XX PR 21-JUL-1999; 99US-0145086.  
XX PR 21-JUL-1999; 99US-0145088.  
XX PR 22-JUL-1999; 99US-0145085.  
XX PR 22-JUL-1999; 99US-0145087.  
XX PR 22-JUL-1999; 99US-0145089.  
XX PR 22-JUL-1999; 99US-0145192.  
XX PR 23-JUL-1999; 99US-0145145.  
XX PR 23-JUL-1999; 99US-0145218.  
XX PR 23-JUL-1999; 99US-0145224.  
XX PR 26-JUL-1999; 99US-0145276.  
XX PR 27-JUL-1999; 99US-0145913.  
XX PR 27-JUL-1999; 99US-0145918.  
XX PR 27-JUL-1999; 99US-0145919.  
XX PR 28-JUL-1999; 99US-0145951.  
XX PR 02-AUG-1999; 99US-0146386.  
XX PR 02-AUG-1999; 99US-0146388.  
XX PR 02-AUG-1999; 99US-0146389.  
XX PR 03-AUG-1999; 99US-0147038.  
XX PR 04-AUG-1999; 99US-0147204.  
XX PR 04-AUG-1999; 99US-0147302.  
XX PR 05-AUG-1999; 99US-0147192.  
XX PR 05-AUG-1999; 99US-0147260.  
XX PR 06-AUG-1999; 99US-0147303.  
XX PR 06-AUG-1999; 99US-0147416.  
XX PR 09-AUG-1999; 99US-0147493.  
XX PR 09-AUG-1999; 99US-0147935.  
XX PR 10-AUG-1999; 99US-0148171.  
XX PR 11-AUG-1999; 99US-0148319.  
XX PR 12-AUG-1999; 99US-0148341.  
XX PR 13-AUG-1999; 99US-0148565.  
XX PR 13-AUG-1999; 99US-0148684.  
XX PR 16-AUG-1999; 99US-0149368.  
XX PR 17-AUG-1999; 99US-0149175.  
XX PR 18-AUG-1999; 99US-0149426.  
XX PR 20-AUG-1999; 99US-0149722.  
XX PR 20-AUG-1999; 99US-0149723.  
XX PR 20-AUG-1999; 99US-0149929.  
XX PR 23-AUG-1999; 99US-0149902.  
XX PR 23-AUG-1999; 99US-0149930.  
XX PR 25-AUG-1999; 99US-0150566.  
XX PR 26-AUG-1999; 99US-0150884.  
XX PR 27-AUG-1999; 99US-0151065.  
XX PR 27-AUG-1999; 99US-0151066.

```

PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 08-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159429.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AAC38511/rev ..

Align seg 1/1 to reverse of: AAC38511 from: 1 to: 1143

666 GlnSerSerLeuGlyProArgSer 673
|||||
467 CAATCCTCTCGGTCGCAAGGAGT 444

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:AAQ38956
seq_documentation_block:
ID AAQ38956 standard; DNR; 1637 BP.
XX AC AAQ38956;
XX DT 28-JUL-1993 (first entry)
XX

```

```

DE Lipase gene #2.
XX
KW Lipase; Pseudomonas; mutant; stability; specificity; ss.
XX
OS Pseudomonas nov. sp. No.109 FERM 3025.
XX
FH Key Location/Qualifiers
FT CDS 13..1380
FT /*tag= a
XX
PN JP05056786-A.
XX
XX 09-MAR-1993.
XX
XX 26-AUG-1991; 91JP-0296730.
XX
XX 26-AUG-1991; 91JP-0296730.
XX
XX (NAGA-) NAGASE SEIKAGAKU KOGYO KK.
XX
XX WPI; 1993-120382/15.
XX
XX P-PSDB; AAR35205.
XX
XX DNA fragment contg. lipase producing gene from Pseudomonas - used
XX to produce large amt. of mutated lipase with increased stability
XX
XX Claim 1; Page 8-11; 13pp; Japanese.
XX
XX The sequences given in AAQ38955-56 encode lipase producing genes from
XX a Pseudomonas species, pref. P. nov. sp. No. 109 FERM 3025. These
XX genes can be used to produce large amount of lipase. They can also be
XX used to produce a mutated lipase which has increased stability or a
XX change in specificity.
XX
XX Sequence 1637 BP; 276 A; 575 C; 539 G; 247 T; 0 other;

alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AAQ38956 ..

Align seg 1/1 to: AAQ38956 from: 1 to: 1637

376 ArgGlnLeuGlyAlaLeuAla 383
|||||
1531 CGACAGCTTGAGGCGCGTGGCG 1554

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH74178
seq_documentation_block:
ID AAH74178 standard; CDNA; 1928 BP.
XX
XX AC AAH74178;
XX
XX 15-OCT-2001 (first entry)
XX
XX Rho family active site of GTPase coding sequence.
DE
XX Rho family active site; GTPase; gene therapy; cancer; haemopathy;
KW HIV infection; immunological disease; inflammation; ss.
XX
XX Unidentified.
XX
XX WO200155420-A1.
XX
XX 02-AUG-2001.
XX
XX 15-JAN-2001; 2001WO-CN00047.
XX

```

PR 26-JAN-2000; 2000CN-01111502.  
 XX (BIOD-) BIODOR GENE TECHNOLOGY LTD SHANGHAI.  
 XX  
 XX  
 PT Mao Y, Xie Y;  
 XX  
 XX WPI; 2001-483260/52.  
 DR P-PSDB; AAG65047.  
 XX  
 XX New Rho family active site of GTPase for diagnosing and treating  
 PT malignant tumor, hemopathy, human immunodeficiency virus infection,  
 PT immunological diseases and various inflammations -  
 XX  
 PS Claim 6; Page 29-30; 37pp; Chinese.  
 XX  
 XX The present invention provides the protein and coding sequence of the Rho  
 CC family active site of GTPase. The sequences can be used in the treatment  
 CC of cancer, haemopathy, HIV infection, immunological diseases and  
 CC inflammation. The present sequence is the coding sequence of the  
 CC invention.  
 XX  
 XX Sequence 1928 BP; 660 A; 370 C; 404 G; 494 T; 0 other;  
 SQ  
 alignment\_scores:  
 Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-09-701-711-2 x AAH74178 ..  
 Align seg 1/1 to: AAH74178 from: 1 to: 1928  
 488 GlyValSerGlnSerLeuSerGly 495  
 821 GGAGTATCTCAAGCCCTTCAGGC 844  
 seq\_name: /SIDS1/gcgdata/geneseq/geneseq-embl/NA2000.DAT:AAH74178  
 seq\_documentation\_block:  
 ID AAC78114 standard; cDNA; 1930 BP.  
 XX  
 XX AAC78114;  
 AC  
 XX  
 XX 08-FEB-2001 (first entry)  
 DT  
 XX  
 XX Human cancer associated gene sequence SEQ ID NO:508.  
 DE  
 XX  
 XX Human; cancer associated gene; cancer antigen; detection; cancer;  
 KW diagnosis; cytostatic; proliferative; vulnerable; immunomodulator;  
 KW antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;  
 KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;  
 KW dermatological; neuroprotective; thrombolytic; coagulant; neutropic;  
 KW vasotropic; antipsoriatic; antidiabetic; antidiabetic; neutropic;  
 KW immune disorder; haematopoietic cell disorder; gene therapy; inflammation;  
 KW allergic reaction; graft versus host disease; autoimmune disorder;  
 KW haemostatic; thrombolytic; cardiovascular disease; organ rejection;  
 KW neurological disease; drug screening; ss.  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 XX WO200055350-A1.  
 PN  
 XX  
 XX 21-SEP-2000.  
 PD  
 XX  
 XX 08-MAR-2000; 2000WO-US05882.  
 PF  
 XX  
 XX 12-MAR-1999; 99US-0124270.  
 PR  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX  
 XX Rosen CA, Ruben SM;  
 PI

XX WPI; 2000-587533/55.  
 DR P-PSDB; AAB43905.  
 XX  
 XX Novel isolated nucleic acids comprising sequences encoding peptides  
 PT useful for treating or diagnosing e.g. cancer -  
 PT  
 XX  
 XX Claim 1; Page 1037-1038; 2352pp; English.  
 PS  
 XX  
 XX AAC77607 to AAC78448 encode the human cancer associated proteins given  
 CC in AAB43398 to AAB44239. The proteins can have activities based on the  
 CC tissues and cells the genes are expressed in. Example of activities  
 CC include: cytostatic; proliferative; vulnerable; immunomodulator;  
 CC antidiabetic; antiasthmatic; antirheumatic; antiarthritic;  
 CC antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;  
 CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;  
 CC neutropic; vasotropic; antipsoriatic and antiangiogenic. The  
 CC polynucleotides and polypeptides can be used for preventing, treating or  
 CC ameliorating medical conditions and diagnosing pathological conditions.  
 CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from  
 CC the present invention may be used to treat immune disorders by activating  
 CC or inhibiting the proliferation, differentiation or mobilisation of  
 CC immune cells, to treat disorders of haematopoietic cells, autoimmune  
 CC disorders, allergic reactions, graft versus host disease and organ  
 CC rejection, modulate haemostatic or thrombolytic activity, modulate  
 CC inflammation, cancers, cardiovascular disorders, neurological disease and  
 CC bacterial or viral infections. The peptides, nucleotides, antibodies,  
 CC agonists and antagonists may be also be used in drug screens. AAC78449 to  
 CC AAC78457 and AAB44240 represent sequences used in the exemplification of  
 CC the present invention.  
 XX  
 XX Sequence 1930 BP; 634 A; 291 C; 320 G; 682 T; 3 other;  
 SQ  
 alignment\_scores:  
 Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-09-701-711-2 x AAC78114/rev ..  
 Align seg 1/1 to reverse of: AAC78114 from: 1 to: 1930  
 703 LeuLeuLeuProLeuProPhelys 710  
 1719 CTTATCTTCACCTCCCTTTAAA 1696  
 seq\_name: /SIDS1/gcgdata/geneseq/geneseq-embl/NA2001B.DAT:ABL05175  
 seq\_documentation\_block:  
 ID ABL05175 standard; cDNA; 2009 BP.  
 XX  
 XX ABL05175;  
 AC  
 XX  
 XX 26-MAR-2002 (first entry)  
 DT  
 XX  
 XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 10007.  
 DE  
 XX  
 XX Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ss.  
 KW  
 XX  
 XX Drosophila melanogaster.  
 OS  
 XX  
 XX WO200171042-A2.  
 PN  
 XX  
 XX 27-SEP-2001.  
 PD  
 XX  
 XX 23-MAR-2001; 2001WO-US09231.  
 PF  
 XX  
 XX 23-MAR-2000; 2000US-191637P.  
 PR  
 XX  
 XX 11-JUL-2000; 2000US-0614150.  
 PI

```
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB61072.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 10007; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2009 BP; 647 A; 459 C; 414 G; 489 T; 0 other;

alignment_scores:
  Quality:      8.00      Length:      8
  Ratio:        1.000     Gaps:      0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x ABL05175/rev ..
Align seg 1/1 to reverse of: ABL05175 from: 1 to: 2009

525 IleAspGluAsnGlnArgIleSer 532
|||||
64 ATCGATGAAACCAACGATTCG 41

seq_name: /SIDSl/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT: AAS81841
seq_documentation_block:
ID AAS81841 standard; cDNA; 2057 BP.
XX
AC AAS81841;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #17645.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA
XX Drmanac RT, Liu C, Tang YT;
PI
XX WPI; 2001-639362/73.
DR
DR P-PSDB; ABG17654.
XX

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID NO 17645; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2057 BP; 505 A; 495 C; 550 G; 507 T; 0 other;

alignment_scores:
  Quality:      8.00      Length:      8
  Ratio:        1.000     Gaps:      0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AAS81841 ..
Align seg 1/1 to: AAS81841 from: 1 to: 2057

370 ValLeuArgArgGluMetArgGln 377
|||||
1231 GTCTCGCTCGGAAATCGTCAG 1254

seq_name: /SIDSl/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT: ABL24790
seq_documentation_block:
ID ABL24790 standard; DNA; 2341 BP.
XX
AC ABL24790;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 25843.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
```

```
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Claim 1; SEQ ID NO 25843; 2lpp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 2341 BP; 616 A; 449 C; 549 G; 727 T; 0 other;
SQ
alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-701-711-2 x ABL24790/rev ..
Align seg 1/1 to reverse of: ABL24790 from: 1 to: 2341
696 LeuAlaThrPheGlySerGluLeu 703
654 TTAGCAACATTTGGCAGCGAGCTA 631
seq_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT: AAD07053
seq_documentation_block:
ID AAD07053 standard; cDNA; 2366 BP.
XX
XX AAD07053;
XX
XX 06-AUG-2001 (first entry)
XX
XX Human DNA-C-MG-72-1776 cDNA clone encoding PRO-C-MG.72 protein.
XX
XX Human; gene therapy; angiogenesis; vascularisation; cardiac hypertrophy;
XX congestive heart failure; CHF; myocardial infarction; aortic stenosis;
XX valvular regurgitation; vascular trauma; wound; burn; cancer; vaccine;
XX endothelial disease; DNA-C-MG.72-1776 cDNA; PRO-C-MG.72 protein; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 71..2062
XX /*tag= a
XX /product= "Human PRO-C-MG.72 protein"
XX
XX WO200125433-A2.
XX
XX 12-APR-2001.
XX
XX 05-OCT-2000; 2000WO-US27512.
XX
XX 07-OCT-1999; 99US-0158587.
XX
XX 28-OCT-1999; 99US-0162611.
XX
XX (GETH ) GENENTECH INC.
XX
```

```
PA (CURA-) CURAGEN CORP.
XX
XX Gerritsen ME, Goddard A, Grimaldi JC, Mehraban F;
PI WPI; 2001-367229/38.
DR P-PSDB; AAE02777.
XX
XX Polypeptides critical for angiogenesis and vascularization, and the
PT nucleic acids encoding them, useful for treating conditions related to
PT inappropriate vascularization and angiogenesis -
XX
XX Claim 10; Page 172-175; 189pp; English.
XX
XX The patent discloses novel polypeptides critical for angiogenesis and
CC vascularisation and nucleic acids encoding them. The sequences of the
CC present invention are used in the prevention, treatment and diagnosis
CC of diseases associated with inappropriate angiogenesis and/or neo- or
CC cardio-vascularisation. They are used in the diagnosis, prevention
CC or treatment of angio-genic disorders such as cardiac hypertrophy,
CC congestive heart failure (CHF), myocardial infarction, aortic stenosis,
CC valvular regurgitation and vascular trauma such as wounds, burns or
CC surgery. They are also used in the treatment of cancer and endothelial
CC diseases. The proteins of the invention are used as antigens in the
CC production of antibodies against the angiogenesis and vascularisation
CC proteins and in assays to identify modulators (agonists and antagonists)
CC of their expression and activity. The antibodies and antagonists are
CC used to down regulate expression and activity of the protein of the
CC invention. The antibodies are also used as diagnostic agents for
CC detecting the presence of the proteins of the invention in samples.
CC The sequences of the invention are also used as vaccines and in gene
CC therapy. The present sequence is human DNA-C-MG.72-1776 cDNA clone
CC encoding novel PRO-C-MG.72 protein of the invention.
XX
XX Sequence 2366 BP; 788 A; 475 C; 539 G; 563 T; 1 other;
SQ
alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-701-711-2 x AAD07053 ..
Align seg 1/1 to: AAD07053 from: 1 to: 2366
488 GlyValSerGlnSerLeuSerGly 495
1321 GGAGTATCTCAAAGCCTTTTCAGGC 1344
seq_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT: AAS94303
seq_documentation_block:
ID AAS94303 standard; cDNA; 2488 BP.
XX
XX AAS94303;
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #30107.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
```

PR 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HYSEQ INC.  
XX Drmanac RT, Liu C, Tang YT;  
XX WPI; 2001-639362/73.  
DR P-PSDB; ABG30116.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
XX Claim 1; SEQ ID No 30107; 103pp; English.  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 2488 BP; 617 A; 677 C; 671 G; 523 T; 0 other;  
  
alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
  
alignment\_block:  
US-09-701-711-2 x AAS94303 ..  
  
Align seg 1/1 to: AAS94303 from: 1 to: 2488  
  
370 ValLeuArgArgGluMetArgGln 377  
|||||  
235 GTCTGCGTCGCGAATGCGTCAG 258  
  
seq\_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.AAS27393  
  
seq\_documentation\_block:  
ID AAS27393 standard; cDNA; 2493 BP.  
XX  
AC AAS27393;  
XX  
DT 07-NOV-2001 (first entry)  
XX  
DE cDNA encoding novel signal transduction pathway protein, Seq ID 428.  
XX  
KW Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;  
KW antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;  
KW immune system disorder; rheumatoid arthritis; inflammatory condition;  
KW organ transplant rejection; infection; hepatitis C; blood disorder;  
KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;  
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;

KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;  
KW reproductive system; gastrointestinal; liver disorder; AIDS; ss;  
KW acquired immune deficiency syndrome.  
OS Homo sapiens.  
XX  
PN WO200154733-A1.  
XX  
PD 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US01312.  
PF  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 05-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.

```
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0251989.
PR 06-DEC-2000; 2000US-0256719.
PR 08-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251899.
PR 08-DEC-2000; 2000US-0251990.

PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465460/50.
DR P-PSDB; AAU17476.
XX
PT Novel polypeptides useful for diagnosing, treating, preventing and/or
PT prognosing disorders related to the proteins, including cancers, immune
PT disorders and neuronal disorders -
XX
PS Claim 1; SEQ ID No 428; 880pp; English.
XX
CC The invention relates to novel isolated polypeptides (I), and
CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
CC diagnosing, preventing and treating diseases including immune system
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections and graft versus host disease, infectious diseases
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC other blood-related disorders (sickle cell anaemia), myeloproliferative
CC disorders, primary haematopoietic disorders, hyperproliferative
CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
CC disorders (e.g. glomerulonephritis), cardiovascular disorders
CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in
CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
CC Addison's disease), reproductive system disorders, gastrointestinal
CC disorder (inflammatory disorders), liver disorders (cirrhosis),
CC as stimulators of B-cell responsiveness to pathogens, activators of
CC T-cells, to induce higher affinity antibodies, and as a means to induce
CC tumour proliferation in pathologies e.g. acquired immune deficiency
CC syndrome (AIDS). AAS26976-AAS27850 represent novel signal transduction
CC pathway protein coding sequences and PCR primers of the invention.
XX

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AAS27393
Align seg 1/1 to: AAS27393 from: 1 to: 2493

488 GlyValSerGlnSerLeuSerGly 495
|||||
309 GGAGTATCTCAAAGCCTTTCAGGC 332
```





OM of: US-09-701-711-2 to: GenEmbl:\* out\_format : pfs  
Date: Sep 19, 2002 1:58 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+p2n.spool -DEV=xlp  
-O=/cgn2\_1/USPTO\_model/US09701711/runat\_17092002\_141401\_26890/app\_query.fasta\_1.882  
-DB=GenEmbl -QPM=fastap -SUFFIX=sepl7std.rge -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blowum62  
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=40 -MODE=LOCAL -OUTPMT=pfs  
-NORM\_ext -HEARSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09701711 -CGN1\_1.8909 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-701-711-2  
Query length: 813  
Database: GenEmbl:\*  
Database sequences: 1797656  
Database length: 187333701  
Search time (sec): 3438.520000

score\_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_pat:AX009287	+ 4202.00	4796.25	8.5e-259	2442	! AX009287 Sequence 1 from Patent
gb_pat:AX009289	+ 4189.00	4781.39	5.7e-258	2442	! AX009289 Sequence 37 from Patent
gb_pat:AX067462	+ 4189.00	4746.15	5.3e-256	99629	! AX067462 Sequence 37 from Patent
gb_pat:AE004784	- 1451.50	1636.86	8.2e-83	12469	! AE004784 Pseudomonas aeruginosa
gb_pat:AJ414146	+ 1285.50	1397.41	1.8e-69	210050	! AJ414146 Yersinia pestis str
gb_pat:AJ236920	+ 1258.50	1426.14	4.5e-71	4405	! AJ236920 Photorhabdus luminesc
gb_pat:AL646064	+ 1237.00	1365.25	1.1e-67	201050	! AL646064 Ralstonia solanacea
gb_pat:AL627266	+ 1230.50	1343.66	1.7e-66	268050	! AL627266 Salmonella enterica
gb_pat:AF407013	+ 1212.00	1377.64	2.2e-68	2700	! AF407013 Escherichia coli oute
gb_pat:AE005193	+ 1212.00	1364.64	1.2e-67	10649	! AE005193 Escherichia coli O15
gb_pat:AE000127	+ 1212.00	1361.42	1.8e-67	14877	! AE000127 Escherichia coli K12
gb_pat:AE008705	+ 1212.00	1347.28	3.0e-67	22997	! AE008705 Salmonella typhimur
gb_pat:ECOTSF	+ 1212.00	1344.17	1.6e-66	91430	! D83536 Escherichia coli genom
gb_pat:ECU070214	+ 1212.00	1341.33	2.4e-66	123171	! U70214 Escherichia coli chro
gb_pat:AP002550	+ 1212.00	1333.48	6.5e-66	281530	! AP002550 Escherichia coli O1
gb_pat:AF120927	+ 1210.00	1376.34	2.6e-68	2433	! AF120927 Shigella flexneri out
gb_pat:AE004297	- 1203.00	1354.09	4.6e-67	10901	! AE004297 Vibrio cholerae chrc
gb_pat:AE003941	- 1172.50	1318.55	4.4e-65	11711	! AE003941 Xylella fastidiosa 9
gb_pat:AX155443	+ 1123.50	1277.68	8.3e-63	2379	! AX155443 Sequence 6 from Patent
gb_pat:AX202493	+ 1123.50	1277.68	8.3e-63	2379	! AX202493 Sequence 6 from Patent
gb_pat:AX155439	+ 1115.00	1267.91	2.9e-62	2394	! AX155439 Sequence 2 from Patent
gb_pat:AX202489	+ 1115.00	1267.91	2.9e-62	2394	! AX202489 Sequence 2 from Patent
gb_pat:AE002375	- 1115.00	1251.44	2.4e-61	13538	! AE002375 Neisseria meningitid
gb_pat:AX043922	- 1115.00	1220.53	1.3e-59	349980	! AX043922 Sequence 1 from Patent
gb_pat:AX155447	+ 1111.00	1263.34	5.2e-62	2394	! AX155447 Sequence 10 from Patent
gb_pat:AX202497	+ 1111.00	1263.34	5.2e-62	2394	! AX202497 Sequence 10 from Patent
gb_pat:NMA12491	+ 1111.00	1216.21	2.0e-59	340806	! AL162752 Neisseria meningitid
gb_pat:AF021245	+ 1109.00	1261.05	7.0e-62	2394	! AF021245 Neisseria meningitid
gb_pat:NGU81959	+ 1104.50	1255.97	1.3e-61	2379	! U81959 Neisseria gonorrhoeae
gb_pat:AF329831	+ 1058.50	1203.38	1.1e-58	2382	! AF329831 Haemophilus ducreyi
gb_pat:AE006235	+ 1054.00	1184.42	1.3e-57	10187	! AE006235 Pasteurella multocid
gb_pat:U32773	- 1053.50	1183.88	1.4e-57	10151	! U32773 Haemophilus influenzae
gb_pat:HIU13961	+ 1052.50	1195.37	3.2e-58	2868	! U13961 Haemophilus influenzae
gb_pat:HIU06833	+ 1052.50	1195.06	3.3e-58	2776	! U60833 Haemophilus influenzae
gb_pat:HIU06834	+ 1052.50	1194.90	3.4e-58	2823	! U60834 Haemophilus influenzae
gb_pat:AX38690	+ 1052.50	1194.49	3.6e-58	2949	! AX38690 Sequence 1 from Patent
gb_pat:AR102148	+ 1052.50	1194.49	3.6e-58	2949	! AR102148 Sequence 1 from Patent
gb_pat:AX38694	+ 1052.50	1194.49	3.6e-58	2950	! AX38694 Sequence 5 from Patent
gb_pat:AR102150	+ 1052.50	1194.49	3.6e-58	2950	! AR102150 Sequence 5 from Patent
gb_pat:AX38692	+ 1052.50	1194.38	3.6e-58	2984	! AX38692 Sequence 3 from Patent

gb\_pat:AR102149 + 1052.50 1194.38 3.6e-58 2984 ! AR102149 Sequence 3 from Patent

gb\_pat:AX38744 + 1052.50 1194.37 3.6e-58 2987 ! AX38744 Sequence 55 from Patent

gb\_pat:AR102157 + 1052.50 1194.37 3.6e-58 2987 ! AR102157 Sequence 55 from Patent

gb\_ba:HIU60832 + 1052.50 1189.03 7.2e-58 5236 ! U60832 Haemophilus influenzae

gb\_ba:HIU60831 + 1050.00 1185.82 1.1e-57 5438 ! U60831 Haemophilus influenzae

seq\_name: gb\_pat:AX009287

seq\_documentation\_block:

LOCUS AX009287 2442 bp DNA linear PAT 06-SEP-2000

DEFINITION Sequence 1 from Patent WO9603093.

ACCESSION AX009287

VERSION AX009287.1 GI:9996618

KEYWORDS

SOURCE Moraxella catarrhalis.

ORGANISM Moraxella catarrhalis

Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;

REFERENCE 1 (bases 1 to 2442)

AUTHORS Basb027 proteins and genes from moraxella catarrhalis, antigens,

TITLE antibodies, and uses

JOURNAL Patent: WO 963093-A 1 09-DEC-1999;

SMITHKLINE BEECHAM BIOLOG (BE); VINALS BASSOLS CARLOTA (BE)

FEATURES

source

1. .2442

/organism="Moraxella catarrhalis"

/db\_xref="taxon:480"

BASE COUNT 737 a 461 c 533 g 711 t

ORIGIN

alignment\_scores:

Quality: 4202.00 Length: 813

Ratio: 5.169 Gaps: 0

Percent Similarity: 100.00 Percent Identity: 100.00

alignment\_block:

US-09-701-711-2 x AX009287

Align seg 1/1 to: AX009287 from: 1 to: 2442

1 MetArgAsnSerTyrPheLysGlyPheGlnValSerAlaMetThrMetAl 17

|||||

1 ATGCGTAATTCATATTTAAAGGTTTTCAGGTCAAGTCAATGCAATGGC 50

|||||

17 aValMetMetValMetSerThrHisAlaGlnAlaAlaAspPheMetAla 34

|||||

51 TGTCTATGATGTAATGTAATCACTCATGCAACAGCGCGGATTTTATGGCAA 100

|||||

34 snAspIleThrIleThrGlyLeuGlnArgValThrIleGluSerLeuGln 50

|||||

101 ATGACATTACCATCACAGGACTACAGGAGTGACCAATGAAAGCTTACAA 150

|||||

51 SerValLeuProPheArgLeuGlyGlnValValSerGluAsnGlnLeuAl 67

|||||

151 AGCGTGTGCGGTTTCGTTGGTGGTCAAGTGTGAGCGAAACCAAGTTGGC 200

|||||

67 aAspGlyValLysAlaLeuTyrAlaThrGlyAsnPheSerAspValGlnV 84

|||||

201 TGTATGGTGTCAAGACACTTTATGCAACAGGCAATTTTCAGATGTGCAAG 250

|||||

84 altYrHisGlnGluGlyArgIleIleTyrGlnValThrGluArgProLeu 100

|||||

251 TCTATCATCAAGAGGCGGTATCATCTATCATAGTAAACCGAAAGCGCGTTA 300

|||||

101 IleAlaGluIleAsnPheGluGlyAsnArgLeuIleProLysGluGlyLe 117

|||||

301 ATCGCTCAGATTATTTTGGAGGCAATCGCTTAATTCARAAGAGGTCT 350

|||||

117 uGlnGluGlyLeuLysAsnAlaGlyLeuAlaValGlyGlnProLeuLysG 134

|||||

351 ACAAGAAGGGCTAAAAAATGCTGGCTTAGCTTGGGTCAACCACTAAAC 400

134 lAlaThrValGlnMetIleGluThrGluLeuThrAsnGlnTyrIleSer 150  
|||||  
401 AGCCACAGTACAGATGATCGAAACCGAGCTTACCAATCAATATATCA 450  
151 GlnGlyTyrAsnThrGluIleThrValLysGlnThrMetLeuAspG1 167  
|||||  
451 CAAGGCTATTATAATACCGAAATTAAGTGTCAACACAGACGATGCTTGATGG 500  
167 yAsnArgValLysLeuAspMetThrPheAlaGluGlyLysProAlaArgV 184  
|||||  
501 TAATCGTGTAAAGTTGATATGACCTTGTCTGAAGGTAAACCTGACGGG 550  
184 alValAspIleAsnIleGlyAsnGlnHisPheSerAspAlaAspLeu 200  
|||||  
551 TGGTTGATATTAAATATCATTTGGCAATCAGCATTTTAGCGATGCGATTG 600  
201 IleAspValLeuAlaIleLysAspAsnLysIleAsnProLeuSerLysAl 217  
|||||  
601 ATTGATGTGCTTGCATTAAGGATAATAAAATCAATCCACTGTCTAAAGC 650  
217 aAspArgTyrThrGlnGlyLysLeuValThrSerLeuGluAsnLeuArgA 234  
|||||  
651 TGACCGTTATACTCAGAAAGCTGTGACCACTTTAGAGATTTCGGTG 700  
234 laLysTyrLeuAsnAlaGlyPheValArgPheGluIleLysAspAlaLys 250  
|||||  
701 CTAATATCTCAATGCAAGGTTTGTGCGTTTTCAGATTAAAGATGCTAAG 750  
251 LeuAsnIleAsnGluAspLysAsnArgIlePheValGluIleSerLeuHi 267  
|||||  
751 CTTAATATTAAAGATATAAAACCCATCTTTGTTGAGATTTCATTGCA 800  
267 sGluGlyGluGlnTyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuT 284  
|||||  
801 TGAAGTGAGCAATATCGCTTTGGACAGACACAGTTTTCGGTAAATTTAA 850  
284 hrTyrThrGlnAlaGluLeuGluAlaLeuLeuLysPheLysAlaGluGlu 300  
|||||  
851 CTTATACTCAAGCAGAACTTGAGGCCTGCTTAAATTCAAAGCAGAAGAA 900  
301 GlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLy 317  
|||||  
901 GGGTTTTCACAGCCATGCTTGAGCAAAACAACAATAATATCAGTACCAA 950  
317 sPheGlyAspAspGlyTyrTyrAlaGlnIleArgProValThrArgI 334  
|||||  
951 ATTTGTCAGATGGCTATTATTATGCTCAATCCCTGCTGAACACGCA 1000  
334 leAsnAspGluSerArgThrValAspValGluTyrTyrIleAspProVal 350  
|||||  
1001 TTAATGATGAAAGTCGTACGGTTGATGTGGAATATTATATTACCCCTGTA 1050  
351 HisProValTyrValArgArgIleAsnPheThrGlyAsnPhelysThrG1 367  
|||||  
1051 CACCCCTGTATGTACGCCGTATTATTTTACAGTAACTTTAAGACCCA 1100  
367 nAspGluValLeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlas 384  
|||||  
1101 AGATGAAGTACTCCCTCGTGAGATGGACAACTTGAAGTGGCTGGCAT 1150  
384 erAsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhe 400  
|||||  
1151 CTAATCAAAAAATCCAGCTGTCTGTCACGCTTGATGCGGACTGGGTTT 1200  
401 PheLysHisValThrValAspThrArgProValProAsnSerProAspG1 417  
|||||  
1201 TTTAAACATGTTTACCGTTGATCTCGTCAGTACCACTTACCCCTGATCA 1250  
417 nValAspValAsnPheValValGluGlnProSerGlySerSerThrI 434  
|||||  
1251 GGTGTGATGTAATTTGTGTTGTCAGAAACAACCTTCAGGATCATCAACA 1300

434 leAlaAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspVal 450  
|||||  
1301 TGCAGCAGGCTACTCTCAAGTGGTGGTAACTTTTCAATTTGATGTT 1350  
451 SerGlnAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSe 467  
|||||  
1351 TCTCAAAATAAAGTTTATGGGTACAGTAAGCAGCTCAATGCTTCGTTTC 1400  
467 rArgSerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrP 484  
|||||  
1401 TCGCTCTGAGACCCGTGAGGTGTATAGTTGGGTATGACCAACCCCATCT 1450  
484 heThrValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThr 500  
|||||  
1451 TTACCGTAAATGGCGTCTCGCAAGCTTGAGTGGCTACTATCGTAAAC 1500  
501 LysTyrAspAsnLysAsnIleSerAsnTyrValLeuAspSerTyrGlyG1 517  
|||||  
1501 AAGTATGATAACAAGAACATTAGTAATTATGATTGATTCTTATGGTGG 1550  
517 ySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSerPheG 534  
|||||  
1551 CTCATTAAAGCTATGGATATCCAAATTCATGATAAATCAACGATAAGCTTG 1600  
534 lyLeuAsnAlaAspAsnThrLysLeuHisGlyGlyArgPheMetGlyIle 550  
|||||  
1601 GTCTGAATGCTGACAAATACCAAGCTTCATGGCGGCTGTTTATGGCATT 1650  
551 SerAsnValLysGlnLeuMetAlaAspGlyGlyLysIleGlnValAspAs 567  
|||||  
1651 AGTAATGTCAAGCAGCTGATGGCAGATGGTGCAAAATTCAGATGGATAA 1700  
567 nAsnGlyIleProAspPheLysHisAspTyrThrThrTyrAsnAlaIleL 584  
|||||  
1701 TAATGCAATTCCTGATTTAAGCATGATTACACACCTACAAATGCCATT 1750  
584 euGlyTrpAsnTyrSerSerLeuAspArgProValPheProThrGlnGly 600  
|||||  
1751 TGGGTGGAATATTATCAAGCTAGATCGCCCTGATTTCACCAACCAAGC 1800  
601 MetSerHisSerValAspLeuThrValGlyPheGlyAspLysThrHisG1 617  
|||||  
1801 ATGAGTCAATTCGTAGATTGACGGTTGGTTTGGTGATAAAACTCATCA 1850  
617 nLysValValTyrGlnGlyAsnIleTyrArgProPheIleLysLysServ 634  
|||||  
1851 AAAAGTGGTTTATCAAGCAATATCTATCGCCCATTTTCAAAAAATCAG 1900  
634 aLeuArgGlyTyrAlaLysLeuGlyTyrGlyAsnAsnLeuProPheTyr 650  
|||||  
1901 TCTTGGTGGATACGCCAAAGTTAGGCTATGGCAATAATTTACCATTTAT 1950  
651 GluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667  
|||||  
1951 GAAAATTTCTATGACGGCGCTATGGTTCCGTTCCGTTGGCTATGATCAATC 2000  
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684  
|||||  
2001 CTCCTTGGGTCCACGCTCACAAAGCTATTTGACAGCTCGCTGGTCAAC 2050  
684 lnThrThrLeuGlyGluValValGlyGlyAsnAlaLeuAlaThrPheGly 700  
|||||  
2051 AAACCACTAGGAGAGGTTTGGTGGTAAATGCTTTGGCAACTTTTCGCG 2100  
701 SerGluLeuIleLeuProLeuProPheLysGlyAspTrpIleAspGlnVa 717  
|||||  
2101 AGTAGCTGATTTTACCTTTGCCATTTAAGGTGATTGGATAGATCAGGT 2150  
717 lArgProValIlePheIleGluGlyGlnValPheAspThrThrGlyM 734  
|||||  
2151 GGTCCAGTGATATTCAATGAGGGCGGTGAGGTTTGTGATACAAACAGGTA 2200  
734 etAspLysGlnThrIleAspLeuThrGlnPheLysAspProGlnAlaThr 750

|||||  
2201 TGGATAACCAACCATGATTTAAACCAATTTAAAGACCCACAAGCAACA 2250  
751 AlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuThrGlnAspLys 767  
|||||  
2251 GCTGAACAAATGCAAGCAGCAATCGCCGCTACTAACCCCAAGATAA 2300  
767 sGlnLeuArgTyrSerAlaGlyValGlyAlaThrTrpTyrThrProIleG 784  
|||||  
2301 ACAGTTGCGTTATAGTCTGCTGCTGGTGTGCAACTTGGTATACGCCCATG 2350  
784 lyProLeuSerIleSerTyrAlaLysProLeuAsnLysLysGlnAsnAsp 800  
|||||  
2351 GTCCTTATCTATTAGCTATGCGCAAGCCATTGAATAAAAAACAAATGAT 2400  
801 GlnThrAspThrValGlnPheGlnIleGlySerValPhe 813  
|||||  
2401 CAGACCGATACGGTACAGTCCAGATTGGTAGTGTCTTT 2439

seq\_name: gb\_pat:AX009289

seq\_documentation\_block:

LOCUS AX009289 2442 bp DNA linear PAT 06-SEP-2000

DEFINITION Sequence 3 from Patent WO9963093.

ACCESSION AX009289

VERSION AX009289.1 GI:9996619

KEYWORDS

SOURCE Moraxella catarrhalis.

ORGANISM Moraxella catarrhalis.

Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;

Moraxella.

REFERENCE 1 (bases 1 to 2442)

AUTHORS Vinals-Bassols,C.

TITLE Basb027 proteins and genes from moraxella catarrhalis, antigens,

antibodies, and uses

JOURNAL Patent: WO 9963093-A 3 09-DEC-1999;

SMITHKLINE BEECHAM BIOLOG (BE); VINALS BASSOLS CARLOTA (BE)

FEATURES

Location/Qualifiers

source

1..2442

/organism="Moraxella catarrhalis"

/db\_xref="taxon:480"

BASE COUNT 735 a 461 c 535 g 711 t

ORIGIN

alignment\_scores:

Quality: 4189.00 Length: 813

Ratio: 5.159 Gaps: 0

Percent Smilarity: 99.877 Percent Identity: 99.754

alignment\_block:

US-09-701-711-2 x AX009289

..

Align seg 1/1 to: AX009289 from: 1 to: 2442

1 MetArgAsnSerTyrPheLysGlyPheGlnValSerAlaMetThrMetAl 17

|||||

1 ATCCGTAATTCATATTTAAAGGTTTCAGTCAATGCAATGCAATGGC 50

|||||

17 aValMetMetValMetSerThrHisAlaGlnAlaAlaAspMetAlaA 34

|||||

51 TGTCATGATGGTAATGCAACTCATGCAAGCGCGGATTTTATGGCAA 100

|||||

34 sNAspIleThrIleThrGlyLeuGlnArgValThrIleGluSerLeuGln 50

|||||

101 ATGACATTCGCCATCAGAGGACTACAGCGAGTGACCATTTGAAGCTTACA 150

|||||

51 SerValLeuProPheArgLeuGlyGlnValValSerGluAsnGlnLeuAl 67

|||||

151 AGCGTCTGCCGTTTGGCTTGGGTCAGTGGTGAGCGAGCACAGTTGGC 200

|||||

67 aSpGlyValLysAlaLeuTyrAlaThrGlyAsnPheSerAspValcInV 84

|||||

|||||

201 TGATGCTGTCAAAGCACTTTATGCACAGGCAATTTTTCAGATGTGCAAG 250  
84 alTyrHisGlnGluGlyArgIleIleTyrGlnValThrGluArgProLeu 100  
|||||  
251 TCATCATCAAGAAGGCGGTATCATCTATCAGGTAAACCGAAGGCCGTTA 300  
101 IleAlaGluIleAsnPheGluGlyAsnArgLeuIleProLysGluGlyLe 117  
|||||  
301 ATCGCTGAGATTAAATTTTGAAGCAATCGCTTAATTCAAAAGAAAGTCT 350  
117 uGlnGluGlyLeuLysAsnAlaGlyLeuAlaValGlyGlnProLeuLysG 134  
|||||  
351 ACAAGAGGGCTAAAAAATGCTGGCTTAGCTGGGTCAACCACTAAAC 400  
134 InAlaThrValGlnMetIleGluThrGluLeuThrAsnGlnTyrIleSer 150  
|||||  
401 AGCCACAGTACAGATGATCGAAACCGAGCTTACCAATCAATATATATCA 450  
151 GlnGlyTyrTyrAsnThrGluIleThrValLysGlnThrMetLeuAspGl 167  
|||||  
451 CAAGGCTATTATAATACCGAAATTTACTGTCAAACAGACGATGCTTGATGG 500  
167 YAsnArgValLysLeuAspMetThrPheAlaGluGlyLysProAlaArgV 184  
|||||  
501 TAATCGTGTAAAGCTTGATATGACCTTTGCTGAAGGTAAACCTGCACGG 550  
184 alValAspIleAsnIleIleGlyAsnGlnHisPheSerAspAlaAspLeu 200  
|||||  
551 TGGTGTATTAATATATCATTTGGCAATCAGCATTTTACGATGCAGATTG 600  
201 IleAspValLeuAlaIleLysAspAsnLysIleAsnProLeuSerLysAl 217  
|||||  
601 ATTGATGCTGCTTCCGATTAAGGATAATAAATCAATCCACTGTCTAAAGC 650  
217 aAspArgTyrThrGlnGluLysLeuValThrSerLeuGluAsnLeuArgA 234  
|||||  
651 TGACCGGTATATCTCAAGAAAGCTGTGACAGCTTTAGAGAAATTTGGGTG 700  
234 laLysTyrLeuAsnAlaGlyPheValArgPheGluIleLysAspAlaLys 250  
|||||  
701 CTAATATCTCAATGCAAGGTTTGTGCGTTTGTGAGATTAAGATGCTAAG 750  
251 LeuAsnIleAsnGluAspLysAsnArgIlePheValGluIleSerLeuHi 267  
|||||  
751 CTTAATATTAAATGAAGATAAAACCGTATCTTGTTCAGATTTCATTGCA 800  
267 sGluGlyGluGlnTyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuT 284  
|||||  
801 TGAAGGTGAGCAATATCGCTTTGGACAGACACAGATTTTGGGTAAATTA 850  
284 hrTyrThrGlnAlaGluLeuGluAlaLeuLeuLysPheLysAlaGluGlu 300  
|||||  
851 CTTATACTCAAGCAGAACTTTGAGGCACTGCTTAAATTCAAAGCAGAGAA 900  
301 GlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLy 317  
|||||  
901 GGGTTTCAACAGCCATGCTTGAGCAACAACAACAATATCAGTACCAA 950  
317 sPheGlyAspAspGlyTyrTyrTyrAlaGlnIleArgProValThrArgI 334  
|||||  
951 ATTTGGTGACGATGGCTATTATTATGCTCAATCCGCTCTGTACACGCA 1000  
334 leAsnAspGluSerArgThrValAspValGluTyrTyrIleAspProVal 350  
|||||  
1001 TTAATGATGAAGTTCGTACCGTTGATGTGGAATATATATTGACCCCTGTA 1050  
351 HisProValTyrValArgArgIleAsnPheThrGlyAsnPheLysThrGl 367  
|||||  
1051 CACCTCTCTATGTACGCCGTATTAAATTTACAGGTAACTTTAAGACCCA 1100  
367 nAspGluValLeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlas 384  
|||||  
1101 AGATGAAGTACTCCGTCGTGAGATGCACAACCTTGAAGGTGCGTTGGCAT 1150

```
384  exAsnGlnIysIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhe 400
1151 CTAATCAAAAATCAGCTGCTCGTGACGCTTGATCGGACTGGGTTT 1200
401  PheLysHisValThrValAspThrArgProValProAsnSerProAspGI 417
1201 TTTAAACATGTTTACCGTTGATACCTCGTCAGTACCAACTCACCTGATCA 1250
417  nValAspValAsnPheValValGluGlnProSerGlySerSerThrI 434
1251 GGTGATGATAAATTTGTGGTTGAGAACAACTTCAGGATCATCAACCA 1300
434  leAlaAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspVal 450
1301 TGCAGCAGCGCTACTCTCAAGTGGTGGTGAACCTTTCAATTTGATGTT 1350
451  SerGlnAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSe 467
1351 TCTCAAAATAACTTTATGGGTACAGTAAGCACGTCATGCTTCGTTTTC 1400
467  rArgSerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrP 484
1401 TCGCTCTGAGACCGCTGAGGTGTATAGTTTGGGTATGACCAACCCATAC 1450
484  heThrValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThr 500
1451 TTACCGTAATAGCGCTCTGCCAAAGCTTGAGTGGCTACTATCGTAAACC 1500
501  LysTyrAspAsnLysAsnIlesSerAsnTyrValLeuAspSerTyrGlyGI 517
1501 AAGTATGATAACAAGAACATTAAGTATATGATGTTGATTTATGGTGG 1550
517  ySerLeuSerTyrGlyTyrProIleAspGlnAsnGlnArgIleSerPheG 534
1551 CCAATTAAGCTATGATATCCATTCATGATGAAATCAACCATCAAGCTTIG 1600
534  lyLeuAsnAlaAspAsnThrLysLeuHisGlyGlyArgPheMetGlyIle 550
1601 GTCTGAATGCTGACAATACCAAGCTTCATGGCGGTGCTTTATGGGCAT 1650
551  SerAsnValLysGlnLeuMetAlaAspGlyGlyLysIleGlnValAspAs 567
1651 AGTAATGTCAAGCAGCTGATGTCAGATGTCGCAAAATCAAGTGGATAA 1700
567  nAsnGlyIleProAspPheLysHisAspTyrThrTyrAsnAlaIleI 584
1701 TAATGGCATTCCTGATTTTAAGCATGATTACACAACCTACATGCCATTT 1750
584  euGlyTrpAsnTyrSerSerLeuAspArgProValPheProThrGlnGly 600
1751 TGGGGTGGAAATATTCAAGTCTAGATCGCCCTGTATTTCACCAACCAAG 1800
601  MetSerHisSerValAspLeuThrValGlyPheGlyAspLysThrHisGI 617
1801 ATGAGTCATCTCTAGATTGACGGTGGTGGTGGTGGTGGTGGTGGTGGT 1850
617  nLysValValTyrGlnGlyAsnIleTyrArgProPheIleLysLysSerV 634
1851 AAAAGTGGTTTATCAAGCAATATCTATCGCCATTTATCAAAAATCAG 1900
634  alLeuArgGlyTyrAlaLysLeuGlyTyrGlyAsnAsnLeuProPheTyr 650
1901 TCTTGGCTGGATACGCAAGTATAGGCTATGGCAATAATTTACCATTTAT 1950
651  GluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667
1951 GAAATTTCTATGACAGCGGCTATGGTTCGGTTCGGTTCGGTTCGGTTC 2000
667  rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
2001 CTCTTTGGGTCCACGCTCACAGCCCTATTGTACAGCTCGTGGTGGTCAAC 2050
```

```
684  InThrThrLeuGlyGluValValGlyGlyAsnAlaLeuAlaThrPheGly 700
2051 AAACACACTAGGAGAGTTGTTGGTGGTAATGCTTTGGCAACTTTTCGGC 2100
701  SerGluLeuIleLeuProLeuProPheLysGlyAspTyrPileAspGlnVa 717
2101 AGTGAGCTGATTTTACCTTTGCCATTTAAAGGTGATGGATAGATCAGGT 2150
717  IargProValIlePheIleGluGlyGlyGlnValPheAspThrThrGlyM 734
2151 GCGTCCAGTATATTCATTCATGAGGCGGTGAGGTTTGTATACAACAGGTA 2200
734  etAspLysGlnThrIleAspLeuThrGlnPheLysAspProGlnAlaThr 750
2201 TGGATAAAACAAACCATTTGATTTAACCAATTTAAAGACCCCAAGCAACA 2250
751  AlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuThrGlnAspLy 767
2251 GCTGAACAAAATGCAAAAGCAGCCAATCGCCGCTACTAACCCCAAGATAA 2300
767  sGlnLeuArgTyrSerAlaGlyValGlyAlaThrTyrTyrThrProIleG 784
2301 ACAGTTGGCTTATAGTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2350
784  lyProLeuSerIleSerTyrAlaLysProLeuAsnLysLysGlnAsnAsp 800
2351 GTCCCTTTATCTATTAGCTATGCCAAGCCATTGAATAAAAAACAATAATG 2400
801  GlnThrAspThrValGlnPheGlnIleGlySerValPhe 813
2401 CAGACCGATACGGTACAGTCCAGATTGGTATGTGTCCTT 2439
seq_name: gb_pat:AX067462
seq_documentation_block:
LOCUS AX067462 99629 bp DNA linear PAT 24-JAN-2001
DEFINITION Sequence 37 from Patent WO0078968.
ACCESSION AX067462
VERSION AX067462.1 GI:12545082
KEYWORDS
SOURCE Moraxella catarrhalis.
ORGANISM Moraxella catarrhalis
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
Moraxella.
REFERENCE
AUTHORS 1 (bases 1 to 99629)
Lagace,R.E., Patterson,C. and Berg,K.L.
TITLE Nucleotide sequences of moraxella catarrhalis genome
JOURNAL Patent: WO 0078968-A 37 28-DEC-2000;
Incyte Genomics, Inc. (US)
FEATURES
source location/Qualifiers
1..99629
/organism="Moraxella catarrhalis"
/db_xref="taxon:480"
78467
/note="a or g or c or t, unknown, or other"
BASE COUNT 29233 a 19222 c 21909 g 29264 t 1 others
ORIGIN
alignment_scores:
Quality: 4189.00 Length: 813
Ratio: 5.159 Gaps: 0
Percent Similarity: 99.877 Percent Identity: 99.754
alignment_block:
US-09-701-711-2 x AX067462
Align seq 1/1 to: AX067462 from: 1 to: 99629
```

```
1 MetArgAsnSerTyrPheLysGlyPheGlnValSerAlaMetThrMetAl 17
ATCGGTAATTCATATTATTTAAAGGTTTTCAGGTCAGTCAATGCAATGGC 66182
```

17 aValMetMetValMetSerThrHisAlaGlnAlaAlaAspPheMetAla 34  
66183 TGTCTATGATGTAATGTCAACTCATGCACAAAGCGCGGATTTATGGCAA 66232  
34 snAspIleThrIleThrGlyLeuGlnArgValThrIleGluSerLeuGln 50  
66233 ATGACATTGCCATCACAGGACTACAGCGAGTGACCAFTGAAGCGTTACAA 66282  
51 SerValLeuProPheArgLeuGlyGlnValValSerGluAsnGlnLeuAl 67  
66283 AGCGTGTCCGTTTCCTTGGGTCAAGTGGTGAGCGAACCACAGTTGGC 66332  
67 aspGlyValLysAlaLeuTyrAlaThrGlyAsnPheSerAspValGln 84  
66333 TGATGGTGTCAAAAGCACTTATGCAACAGGCAATTTTTCAGATGTGCAAG 66382  
84 alTyrHisGlnGluGlyArgIleIleTyrGlnValThrGluArgProLeu 100  
66383 TCTATCATCAAGAAGGCGGTATCATCTATCAGGTAACCGAAGGCGGTTA 66432  
101 IleAlaGluIleAsnPheGluGlyAsnArgLeuIleProLysGluGlyLe 117  
66433 ATCGCTGAGATTAATTTTGAGGCAATCGCTTAATTCCAAAGAAGGCT 66482  
117 uGlnGluGlyLeuLysAsnAlaGlyLeuAlaValGlyGlnProLeuLys 134  
66483 ACAAGAAGGGCTAAAAAATGCTGGCTTAGCTGGGTCAACCACCTAAAC 66532  
134 lnaLanThrValGlnMetIleGluThrGluLeuThrAsnGlnTyrIleSer 150  
66533 AAGCCACAGTACAGATGATCGAAACCGAGCTTACCAATCAATATATATCA 66582  
151 GlnGlyTyrTyrAsnThrGluIleThrValLysGlnThrMetLeuAspG 167  
66583 CAAGGCTATTAATACCGGAATTAATGTCAAACAGACGATGCTTGATGG 66632  
167 YAsnArgValLysLeuAspMetThrPheAlaGluGlyLysProAlaArg 184  
66633 TAATCGTGTAAAGCTTGATATGACCTTTGCTGAAGGTAAACCTGCACGG 66682  
184 alValAspIleAsnIleIleGlyAsnGlnHisPheSerAspAlaAspLeu 200  
66683 TGCTTGATATTAATATCATTTGGCAATCAGCATTTTAGCGATGCAGATTTG 66732  
201 IleAspValLeuAlaIleLysAspAsnLysIleAsnProLeuSerLysAl 217  
66733 ATTGATGTCTCCGATTAAGGATATAAATCAATCCACTGCTAAAGC 66782  
217 aAspArgTyrThrGlnGluLysLeuValThrSerLeuGluAsnLeuArg 234  
66783 TGACCGTTTACTCAAGAAAGCTGGTGACCACTTTAGAGAAATTTGCGTG 66832  
234 laLysTyrLeuAsnAlaGlyPheValArgPheGluIleLysAspAlaLys 250  
66833 CTAATATCTCAATGACGGGTTTGTCCGTTTGGAGATTAAAGATGCTAAG 66882  
251 LeuAsnIleAsnGluAspLysAsnArgIlePheValGluIleSerLeuHi 267  
66883 CTTAATTAATGAAGATAAAAACCGTATCTTTGTTGAGATTTCAATTGCA 66932  
267 sGluGlyGlnTyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuT 284  
66933 TGAAGGTGAGCAATATCGCTTTGGACAGACACAGATTTTGGGTAATTTAA 66982  
284 hrTyrThrGlnAlaGluLeuGluAlaLeuLeuLysPheLysAlaGluGlu 300  
66983 CTTATACTCAAGCAGAACCTTGAGGCACTGCTTAAATTTCAAGCAGAGAA 67032  
301 GlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLy 317  
67033 GGGTTTTCACAGCCATGCTTGAGCAGAACACAAACAATATCATGTACCAA 67082  
317 sPheGlyAspGlyTyrTyrTyrAlaGlnIleArgProValThrArgI 334

67083 ATTTGGTGACGATGGCTATTATTATGCTCAAAATCCGTCCTGTAAACGCA 67132  
334 leAsnAspGluSerArgThrValAspValGluTyrIleAspProVal 350  
67133 TTAATGATGAAGTCGTACGGTTGATGTGGAATATTATATTGACCCCTGTA 67182  
351 HisProValTyrValArgArgIleAsnPheThrGlyAsnPheLysThrGl 367  
67183 CACCCTGCTATGTACGCCGCTATTAAATTTACAGGTAACTTTTAAGACCCA 67232  
367 nAspGluValLeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlas 384  
67233 AGATGAAGTACTCCGTCGTGAGATGCCACAACCTTGAAGGTGCGTTGGCAT 67282  
384 erAsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhe 400  
67283 CTAATCAAAAAATCCAGCTCTCGTGCACGCTTGATGCGGACTGGGTTT 67332  
401 PheLysHisValThrValAspThrArgProValProAsnSerProAspGl 417  
67333 TTTAAACATGTTACCGTTGATACTCTGTCAGTACCCCAACTCACCTGATCA 67382  
417 nValAspValAsnPheValValGluGlnProSerGlySerSerThrI 434  
67383 GGTGATGTAATTTTGTGCTTGAAGAACAACCTTTCAGGATCATCAACCA 67432  
434 leAlaAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspVal 450  
67433 TCCACGACGGCTACTCTCAAAAGTGGTGTGTAACCTTTTCAATTTGATGTT 67482  
451 SerGlnAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSe 467  
67483 TCTCAAAATTAACCTTATGGGTACAGGTAAGCACGTCATGCTTCGTTTTC 67532  
467 rArgSerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrP 484  
67533 TCGCTCTGACACCCGTCGAGGTGATAGTTTGGGTATGACCAACCCATACT 67582  
484 heThrValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThr 500  
67583 TTACCCGTAATGGCGTCTCGCAAGCTTGAGTGGCTACTATCGTAAACCC 67632  
501 LysTyrAspAsnLysAsnIleSerAsnTyrValLeuAspSerTyrGlyGl 517  
67633 AAGTATGATTAACAAGAACATTAATTAATTAATTAATTAATTAATTAATTA 67682  
517 ySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSerPheG 534  
67683 CTCATTTAAGCTATGGATATCCAATTTGATGAAAAATCAACGCATAAGCTTTG 67732  
534 lylLeuAsnAlaAspAsnThrLysLeuHisGlyGlyArgPheMetGlyIle 550  
67733 GTCGTAATGCTGACAAATACCAAGCTTCATGGCGGTGCTTTTATGGGCATT 67782  
551 SerAsnValLysGlnLeuMetAlaAspGlyGlyLysIleGlnValAspAs 567  
67783 AGTAATGTCACAGCAGCTGATGGCAGATGGTGGCAAAAATTCAAAGTGGATA 67832  
567 nAsnGlyIleProAspPheLysHisAspTyrThrThrTyrAsnAlaIle 584  
67833 TAATGGCATTCCTGATTTTAAGCATGATTACACAACCTACAATGCCATTT 67882  
584 euGlyTrpAsnTyrSerSerLeuAspArgProValPheProThrGlnGly 600  
67883 TGGGGTGGAAATTAATCAAGTCTAGATCGCCCTGTATTTTCAACCCCAAGGC 67932  
601 MetSerHisSerValAspLeuThrValGlyPheGlyAspLysThrHisGl 617  
67933 ATGAGTCATTCGTAGATTTGACGGTGGTTTGGTGATAAAACTCATCA 67982  
617 nLysValValTyrGlnGlyAsnIleTyrArgProPheIleLysLysSerV 634

```

67983 AAAAGTGGTTTATCAAGGCAATATCATTCGCCCATTTATCAAAAAATCAG 68032
634 aLeuArgGlyTyrAlaLysLeuGlyTyrGlyAsnLeuProPheTyr 650
|||||
68033 TCTTCGGTGGATACGCCAAGTAGGCTATGCAATAATTTACCATTTAT 68082
|||||
651 GluAsnPheTyrAlaGlyCylTyrGlySerValArgGlyTyrAspGlnSe 667
|||||
68083 GAAAAATTTCTATGACGGCGGCTATGTTGGTTCGGTTCGTATGATCAATC 68132
|||||
667 tSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgGlyGlnG 684
|||||
68133 CTCCTTTGGGTCCACCTCACAGCCTATTGACAGCTCGTGGTGTCAAC 68182
|||||
684 InThrThrLeuGlyGluValValGlyGlyAsnAlaLeuAlaThrPheGly 700
|||||
68183 AAACACACTAGGAGAGGTTGTTGGTGTATGCTTTGGCACTTTCGGC 68232
|||||
701 SerGluLeuLeuLeuProLeuProPheLysGlyAspTrpIleAspGlnVa 717
|||||
68233 AGTAGCTGATTTTACCTTTGCCATTTAAAGGTGATTGATAGATCAGGT 68282
|||||
717 IArgProValIlePheIleGluGlyGlyGlnValPheAspThrThrGlyM 734
|||||
68283 GCGTCCAGTGATATTTCATTAGGCGCGTCAGGTTTTTATACACAGGTA 68332
|||||
734 eAspLysGlnThrIleAspLeuThrGlnPheLysAspProGlnAlaThr 750
|||||
68333 TGGATAAACAACCATTTGATTTAACCCAAATTTAAAGACCCACAACACA 68382
|||||
751 AlaGluInAsnAlaLysAlaAlaAsnArgProLeuLeuThrGlnAspLys 767
|||||
68383 GCTGAACAAAAATGCAAAAGCAGCCCAATCGCCGCTACTTAACCCAAATAA 68432
|||||
767 scLeuArgTyrSerAlaGlyValGlyAlaThrTrpThrProIleG 784
|||||
68433 ACAGTTCGCTATATAGTCTGGTGGTGGTGGCACTTGGTATAGCCCATGTG 68482
|||||
784 lyProLeuSerIleSerTyrAlaLysProLeuAsnLysLysGlnAsnAsp 800
|||||
68483 GTCCTTTATCTATTAGCTATGCAAGCCATTGAATAAAAAACAATAATGAT 68532
|||||
801 GlnThrAspThrValGlnPheGlnIleGlySerValPhe 813
|||||
68533 CAGACCGATACGATGATGATGATGATGATGATGATGATGATGATGATGAT 68571
|||||
seq_name: gb_ba:AE004784

seq_documentation_block:
LOCUS AE004784 12469 bp DNA linear BCT 30-AUG-2000
DEFINITION Pseudomonas aeruginosa PA01, section 345 of the complete
genome.
ACCESSION AE004784 AE004091
VERSION AE004784.1 GI:9949799
KEYWORDS
SOURCE Pseudomonas aeruginosa.
ORGANISM Pseudomonas aeruginosa
Bacteria: Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
REFERENCE
1 (bases 1 to 12469)
Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
Hickey,M.J., Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M.,
Garber,R.L., Goltry,L., Tolentino,E., Westbrook-Wadman,S., Yuan,Y.,
Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,
Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.
Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen
Nature 406 (6799), 959-964 (2000)
JOURNAL MEDLINE
2043737
REFERENCE
2 (bases 1 to 12469)
Stover,C.K., Pham,X.-O.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,

```

```

Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Sailer,M.H.,
Hancock,R.E.W., Lory,S. and Olson,M.V.
Direct Submission
Submitted (16-MAY-2000) Department of Medicine and Genetics,
University of Washington Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
FEATURES
Location/Qualifiers
1..12469
/organism="Pseudomonas aeruginosa"
/strain="PA01"
/db_xref="taxon:287"
complement(106..3627)
/gene="dnaE"
/note="PA3640"
complement(106..3627)
/gene="dnaE"
/codon_start=1
/transl_table=11
/product="DNA polymerase III, alpha chain"
/protein_id="AAG07028.1"
/db_xref="GI:9949800"
/translation="MTYSVHLRLHTEFSLVGLVRVKPLAKAVAGLGMPAVATDQS
NMCSLVFKTAMGAKIPICGADIWLASREEDGPIRSLSLLAMNAKGYRNLTELISR
GWSQGRNGEIIIRSDWVKEAEGLIALSAAKEGEIGHALLDGEAEAEALLKHEWVEV
FPERFYVEQRTSRVNDHEHLAAVALASRCNAPLVATNDVRFIKQDEFAEHFRVCI
GEGRTLDDPRRPTYSDOQLKSPAEMAEFLSDLPALENTEIAKRCNTEVOLGYIF
LPDPTPNGMGIDDLRHASFEGLEERLEVLKPKDTPDYEAQRQVTVDRNLRFELDII
QMGPFGYFLIYMDFIKWKNNVPGVGRSGAGSLVAVYKLTIDLDPLAYDLLEFRF
LNPERSMPDFDVFCEMGDRVDIVYADAYGRNAVSIITFTMAKAVVRDVARVQ
KSYGLADRLSRMPEFVGMTLDKAYEQEEMLRDLKSDSEAAEIWMALKLEITRG
TGHAGGVVIAPTKLTFDSP IACDEEGGLVTFQDKDDVEAGLVKFDLGLRTLII
KWAMEIINREAAKGLPEVNIDFIPLODKPTYSLLOKAETTAVFQESRGMKELIKKL
KPCLEDLIALVALFRPGPLOGSMVDVFNKRGRAELSYPHDPDYAGLEVPVKPTY
GIILYQEQWQIAQWVAGTLCGADMLRRAMKKKPEEMAKQRGEGIEGCKNNGIDAD
LAINFDLVEKFAGYFNKSHSAAYGLVSTQATLAWLHFPAPFMAAVLTADMQNTDKV
VTLIECRHMKLRIVAPVNDSEFRVTDGGRIVYGLGAIKGVGEGPVEAITECHRAE
GGPNTLFQCDVRDLKIRNKRTLEALIRAGALDRGLPHYHDELKAYQATVDLNRVL
LAAMEAIIQAQETARSHDSHMDLFGVFAEPAEDVYANHRVKELTLKRLKGEKD
TLGLYLTGHPIDEEVEGEVRRFARIVELKPARDTQTVAGLIVNLRMKNKGDKNKG
VILDRSGRIEASLFSEAPAAQSLLQTDALVYVEGEVSQDDFSGGLRLRAKRVMSLE
EARTGLAESURMKLHADLLKDRLWLGELFNHRGSCPTITDITTSADAKALLQFGES
WRVDPADDLTIQALRDQFGRDNFNLNR"
/gene="PA3641"
complement(3757..5172)
/gene="PA3641"
complement(3757..5172)
/codon_start=1
/transl_table=11
/product="probable amino acid permease"
/protein_id="AAG07029.1"
/db_xref="GI:9949801"
/translation="MQEFATQVVDLLNGLWGLKILWLLVGCGLYETVRLGLIQFRHF
GHTFVLKGRSDSDSGISFQALCTSLAARGVGNVAGVAITLGGGAVFMMWMAI
ALVGMATGEVETATLQFLIRDDKQGRFGGPAYIMKGLGARWMLFSLFLIAFGF
VNSVQANTITGAMOGAFGPTWISGVIIVVITLAIIFGGRSIAFSELAIVPFAVGA
YLLAVGIIILNLSGLPVLAMVVKSAFGWHEAAAGGIGAAIILNGFKRGLFSNEAGM
SAPNAASATPPYPHPSAGGYQMGAVFIDTLICTASRAIILLAGQEGEGTMLVQN
ALTSQVQDWGRFLVAVIILFFATFSIVANFYAENCLVFLHNLHLAGLLIFRLIVLAM
VMFGALASLPFVNLADVSMGLMAITNLIAILLLSNLAIKAKDYNAQRKAGKLPTFD
ASQFPVQOKLEFGIWDHRA"
complement(5389..5994)
/gene="rnhB"
/note="PA3642"
complement(5389..5994)
/gene="rnhB"
/codon_start=1
/transl_table=11
/product="ribonuclease HII"
/protein_id="AAG07030.1"
/db_xref="GI:9949802"
/translation="MQLGDLFNLVEDLVAGVDEVGKPGICGPPVVTAAVILDPSPILG

```

```

LNDKSKLSEARREALFEETREKALAWCIARAEEVIDRLNIIHLMLAMORAVEGLSV
IPRLIDGNRCPKIAPVPCAPVVGKDSOVPATAAASILAKYSRDRMVELDRLVPIYCYG
MAGHKGYTPAVHLEALSRLGTPPIHRRSFAPVRELLDVSVQ"
/gene="lpxB"
/complement(5994. .7130)
/gene="lpxB"
/complement(5994. .7130)
/codon_start=1
/transl_table=11
/product="lipid A-disaccharide synthase"
/protein_id="AAG07031.1"
/db_xref="GI:9949803"
/translation="MADGLRVALVAGEASGDTLGSGLMOALRARHPDIEFTICVGGPRM
EAGLSSYFPERLSVGLVEVLGRLPELLRRKRLILIEARPDVIMIGIDAPDFTL
GVEHKLROGLRTHVYSPSWAMQKRVKIRACDMLALFPEARFYEEHGVFVR
FVGHPLANTIPQADARAARLGLPADQVIALMPGSGGVEALCALDLDIAQORLL
VPEGLRVLPCASAREQIEQMLQGREPLPLTLDDGSHAEALAACDAVLIASTGAT
LEALYLRPMVAVRVAGLTYRILKRLVKSYPISLPLLALGRLLVPELIQDAATPOAL
AATLSPLDDGSOQVEFFDAIHRALRQDASQAQAEAVLQVERR"
/complement(7134. .7910)
/gene="lpxA"
/complement(7134. .7910)
/gene="lpxA"
/complement(7134. .7910)
/codon_start=1
/transl_table=11
/product="UDP-N-acetylglucosamine acyltransferase"
/protein_id="AAG07032.1"
/db_xref="GI:9949804"
/translation="MSLIDPRAIIPDSARLAADVGVGWSIVGAEEIGEVTIGVPHV
VLKQPTKIGKNNRIQVSSVEDTDLAYKGEPTRLVIGDHNVIREGVYIHRGTWQDR
AETTGIDNLIIMAYAHIGHDSVIGNHCILVNNALAGHVHDDWAILSGYTLVHOYCR
IGANSFSGMGAIGKDPAYVTVFQNPAAEKSMFNGRRRFGFSEALHAKRATKYV
YRQGTVEEALAEASQAQFPAVAVFRDTSQSATRGITR"
/complement(7907. .8347)
/gene="fab2"
/complement(7907. .8347)
/gene="fab2"
/complement(7907. .8347)
/codon_start=1
/transl_table=11
/product="3(R)-hydroxymyristoyl-{acyl carrier protein}"
/protein_id="AAG07033.1"
/db_xref="GI:9949805"
/translation="MMDINEIREYLPHPYPELLVDVVRVELDIEGKRIRAYKNVSNIEP
FNGHFPHPHMPGVLLIEAQAAGILGFRLMDVKPADGLIYFVGSKLRFROPVL
PDQQLHAKFISVKRSIWKFDCHATVDKPKVCSAEIICAERKL"
/complement(8393. .9454)
/gene="lpxD"
/complement(8393. .9454)
/gene="lpxD"
/complement(8393. .9454)
/codon_start=1
/transl_table=11
/product="UDP-3-O-[3-hydroxylauroyl]"
/protein_id="AAG07034.1"
/db_xref="GI:9949806"
/translation="MMSTLSVTLGQLAHVGAENVGDDADLPLOGLATLQEAQPAQLSF
LANPQYRKLPESRAGAVLLAADRGAGTALVVANPILAYASLSHLDFDRKPKAAAG
IHPTAIAADAEVDSVASGAYAVIESGAFAGVIGAHCVIGARSYVIGEGWLAAPR
VTLVHDVYIGARVTSQGVIGGEGFANFKGVQKIAQIGGVYIGDDVEIGANTTI
DRGALSDFLIGNVKVLDNIMIAHNVQIGDHTMAACVIGISAKIGRHCMLAGGVGL
VGHTEICDNVFTGTMVYTRSTITEPGSYSSGTAMPQPAEENKKSARIQLDDMARLQ
QLEKRLAATVSSGDASSDA"
/complement(9454. .9960)
/gene="PA3647"
/complement(9454. .9960)
/gene="PA3647"
/codon_start=1
/transl_table=11
/product="probable outer membrane protein precursor"
/protein_id="AAG07035.1"

```

```

/db_xref="GI:9949807"
/translation="MRKTFQVILITAAIIPAASFAEMKIAVLNOMALLESDAAKQY
AVDAEKKFGPQLNKLNLERAKADQLKLVNKGSKMSQDREKAEIDFKQARDFOFQV
SKELNESKAAADRDLKKLPKLDQAVEETIRKGGYDMVIERGAVVDRPQYDITRQV
TERMNQLR"
/complement(10011. .12404)
/gene="PA3648"
/complement(10011. .12404)
/gene="PA3648"
/codon_start=1
/transl_table=11
/product="probable outer membrane protein"
/protein_id="AAG07036.1"
/db_xref="GI:9949808"
/translation="MKRELLPALLSALMTAEVAESFTVSDIRVNLQVRSAGSVFAA
LPNLVGETIDQALVQATKSLFKTFQDIQROGNDVLTVTVERPSISSIEIEGNK
AISKEDLLKQSGLAIEIFQRATLEGRNELQVYVAQGRYSAEIAEVIPOPRN
RVALKINNEGTVAAISHINVTQVSEEDITDLFELKTNLWFFKNDDKYAREKL
SGDLERLRYLLDRGYINMDIASTQVSTPDKKHVITVINEGEKYTTRDKVLTGDL
KVPEEVRLLLVQKQVFSRKVMTTSDLTRRLNGEYTFANYNGVPEAHDDDKTV
SVTFVDFGKRAYVNRINFRGNTKTEDEVLRERMQMEGWSASTYLIDQSKARLERLG
YFKEVNETPAVPTDDQVDVNVISVEEQSGSITASVGAQAGLILGSI SONNFI
TGKNSIGLITREYQTYRNFQVDPYTWVDSGLYNAFYRKTDDDELVDVVASYSVN
SLGAGMSIGYPISETSRITGLSVQDQIDTGRYTVDEIYDFLDKEDGNTFKASIG
WSESTLNKGLVATRGHSQSLTLETLPGSGLSFKYIDYRGQVFAPLTDNKTFRFTEL
GYGDGYSSTERLPFVYKAGFNSVRGPKDSTLGRPSPSVARNPDGTPMKNOGPD
KGRYTDPDQDPEAFGNNILITGGABELLPPLPFVKDQORLRTVLFWDVSGSTFDTPK
TTTNCDDGKIDNLASSVGLTWITALGPLSLATPIKPPDNAETQVQFSLGQTF"
BASE COUNT 2144 a 3895 c 4089 g 2341 t
ORIGIN
alignment_scores:
Quality: 1451.50 Length: 835
Ratio: 2.460 Gaps: 13
Percent Similarity: 70.659 Percent Identity: 36.886
alignment_block:
US-09-701-711-2 x AE004784/rev ..
Align seg 1/1 to reverse of: AE004784 from: 1 to: 12469
7 LysGlyPheGlnValSerAlaMetThrMetAlaValMetMetValMetSe 23
12401 AAAGCGTTCTGCTACCGCGCTG...CTGTCAGCGTGTATGATCGCGCA 12355
23 rThrHisAlaGlnAlaAlaaspPheMetAlaAsnAspIleThrIleThrG 40
12354 GGTTCACGCGAGTCC...TTCACTGTTCCGATATCGGGTCAATG 12311
40 lYLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPheArg 56
12310 GCCTGCAGCGGTGTCGCGCGTAGCGTGTTCGCGCGCTGCGCGTGAAC 12361
57 LeuGlyGlnValValSerGluAsnGlnLeuAlaAspGlyValLysAlaLe 73
12260 GTCCGGCAGACCATCGACACCGCCCTGTGTCAGGCCACCGCCCTCCCT 12211
73 uTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisGlnGluGlyA 90
12210 GTTCAAGACCGGCTTCTCCAGGACATCCAGCTCGCGCGGTGATGGCAACG 12161
90 rGlleIleTyrGlnValThrGluArgProLeuIleAlaGluIleAsnPhe 106
12160 TGTGGTAGTCACTGTAGTCGAGCGCGCTCGATCTCCAGCATCGAGATC 12111
107 GluGlyAsnArgLeuIleProLysGluGlyGlnGluGlyLeuLysAs 123
12110 GAGGGCAACAGGCGATCTCCAAGAAACCTGCTCAAGGGCCCTGAAACA 12061
123 nAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrValGlnMetI 140
12060 GTCGGCGCTGCGCGAGGCGAGATCTTCCAGCGCGCCACCCTCGAAGCG 12011

```



```
140 leGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyrTyrAsnThr 156
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
12010 TCGCAACAGAGCTGCAACGCCAGCTAGCTGGCCCGCCAGGCGCTACTCGGCC 11961
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
157 GluIleThrValLysGlnThrMetLeuAspGlyAsnArgValLysLeuAs 173
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
11960 GAGATCAACCGCGAGTATCCCGCAGCGCGCAACCGCGCTCGCGTGAA 11911
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
173 pMetThrPheAlaGluGlyLysProAlaArgValValAspIleAsnIleI 190
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
11910 GATCAACATCAACGAGGACCGCTGCCCGGATTTCCCATCATCAACGTGG 11861
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
190 leGlyAsnGlnHisPheSerAspAlaAspLeuLeuValLeuAlaIle 206
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
11860 TGGGCAACACCGCTTCTCCGAAGAAGACCTGACCGACCTGTTCGAACATG 11811
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
207 Lys...AspAsnLysIleAsnProLeuSerLysAlaAspArgTyrThrG1 222
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
11810 AGACCACCAACTGGCTGTCTTCAAGAAGACGACGACAAGTACGCCCG 11761
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
222 nGluLysLeuValThrSerLeuGluAsnLeuArgAlaLysTyrLeuAsnA 239
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
11760 CGAAAGCTCTCCGGCAGCTCGAGCGCTCGCTCTCTATTACTTGAGACC 11711
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
239 laGlyPheValArgPheGluIleLysAspAlaLysLeuAsnIleAsnGlu 255
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
11710 GCGGCTACATCAACATGGATATCGCTCCACCCAGGTATCCATCACCCCG 11661
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
256 AspLysAsnArgIlePheValGluIleSerLeuHisGluGlyGluGlnTy 272
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
11660 GACAAGAAGCAGCTGACATCAACCGCTCAACATCAACGAGGTGAGAAGTA 11611
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
272 rArgPheGlyGlnThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAlaAg 289
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
11610 CACCATCCGGCAGTGAAGTACCGCGGACCTGGAAGGTCCCGGAAGAGG 11561
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
289 luLeuGluAlaLeuLeuLysPheLysAlaGluGlyPheSerGlnAla 305
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
11560 AAGTGAAGCGCTGCTGTGTACAGAAGGCCAGGTGTCTCGCGCAAG 11511
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
306 MetLeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAspG1 322
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
11510 GTGATGACCACCACTCCGACCTTAATCACCCGCTCGCCTGGGCAACGAGG 11461
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
322 yTyrTyrAlaGlnIleArgProValThrArgIleAsnAspGluSerA 339
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
11460 CTACACCTTCGCCAACGTCAACGCGGTGCGGAAGCCCATGACGAGACA 11411
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
339 rgThrValAspValGluTyrTyrIleAspProValHisProValTyrVal 355
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
11410 AGACCGTCTCGGTGACCTTCGTCTGCGACCGCGCAAGCGCGCTACGTC 11361
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
356 ArgArgIleAsnPheThrGlyAsnPheLysThrGlnAspGluValLeuAr 372
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
11360 AACCGCATCAACTTCGCGGCACACCAAGACCGCAAGACGAAGTGTGCG 11311
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
372 gArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLysIleG 389
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
11310 CCGGAATCGCCAGATGGAAGCGGCTGGGCGCTCCACCTACCTGATCG 11261
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
389 lnLeuSerArgAlaArgLeuMetArgThrGlyPhePheLysHisValThr 405
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
11260 ACCAGTCAAGGCGGCTCTGGAGCGCTCGGCTACTTCAAGGAAGTCAAC 11211
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
406 ValAspThrArgProValProAsnSerProAspGlnValAspValAsnPh 422
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
11210 GTGAGACCCCGCGGTTCGCGCAGCCGACGACGAGGTGACGCTCAACTA 11161
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
422 eValValGluGlnProSerGlySerSerThrIleAlaAlaGlyTyrS 439
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
11160 CAGCGTGAAGAGACAACCGTCCGCTCGATCACCGCGCAGCGCTCGGCTCG 11111
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
439 erGlnSerGlyGlyValThrPheGlnPheAspValSerGlnAsnAsnPhe 455
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```



10255 ACCAG...CGCCAGTTCGCTACCGTCTGCTTCTGGGACGTGGGTAGCAC 10209

729 PheAspThrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPheLy 745

10208 TTGATACCGACTGCGCGACACAGACACCAACCACTGCGACGCATCAA 10159

745 sasPProGlnAlaThrAlaGluGlnAsnAlaLysAlaAAsnArgProL 762

10158 G..... 10158

762 euLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThr 778

10157 .....ACCGACAACTCGCCACTTCGCTGGCGCGTCCGCGTAC 10119

779 TrpTyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeuAs 795

10118 TGCATCACCGCGTGGCGCGCTGAGCTTCAGCTGCGGACGCGATCAA 10069

795 nLysLysGlnAspGlnThrAspThrValGlnPheGlnIleGlySerV 812

10068 GAAGCGGACACGCGGAAACCGAGTCTTCCAGTCTCCCTGGGCCAGA 10019

812 alPhe 813

10018 CCTTC 10014

seq\_name: gb\_ba:AJ414146

seq\_documentation\_block: 210050 bp DNA linear BCT 04-OCT-2001  
LOCUS AJ414146  
DEFINITION Yersinia pestis strain C092 complete genome; segment 6/20.  
ACCESSION AJ414146 AL590842  
VERSION AJ414146.1 GI:15979072

KEYWORDS Yersinia pestis.

SOURCE Yersinia pestis.

ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

Yersinia.

REFERENCE 1 (bases 1 to 210050)  
AUTHORS Parkhill,J., Wren,B.W., Thomson,N.R., Titball,R.W., Holden,M.T.G.,  
Prentice,M.B., Sebahia,M., James,K.D., Churcher,C., Mungall,K.L.,  
Baker,S., Basham,D., Bentley,S.D., Brooks,K., Cerdeno-Tarraga,A.M.,  
Chillingworth,T., Cronin,A., Davies,R.M., Davis,P., Dougan,G.,  
Feltwell,P., Hamlin,N., Holroyd,S., Jagels,K., Leather,S.,  
Karllyshev,A.V., Moule,S., Oyston,P.C.F., Quail,M., Rutherford,K.,  
Simmonds,M., Skelton,J., Stevens,K., Whitehead,S. and Barrell,B.G.  
Genome sequence of Yersinia pestis, the causative agent of plague  
Nature 413 (6855), 523-527 (2001)

21470413

REFERENCE 2 (bases 1 to 210050)

Parkhill,J.

Direct Submission

TITLE Submitted (04-OCT-2001) Submitted on behalf of the Yersinia

sequencing team, Sanger Centre, Wellcome Trust Genome Campus,

Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk

COMMENTS

Details of Y. pestis sequencing at the Sanger Centre are available

on the World Wide Web.

(URL, http://www.sanger.ac.uk/Projects/Y\_pestis/).

FEATURES

source

1. .210050

/organism="Yersinia pestis"

/strain="C092"

/db\_xref="taxon:632"

/note="biovar: Orientalis"

829. .11946

/gene="yapH"

829. .11946

/gene="yapH"

/note="Previously sequenced as Yersinia pestis yapH

protein TR:CA14227 (EMBL:AJ277631) (3705 aa) fasta

scores: E(): 0, 99.6% id in 3705 aa. Similar in regions to

Escherichia coli Ycha protein TR:09JMS3 (EMBL:AP001918)  
(1371 aa) fasta scores: E(): 0, 39.1% id in 1405 aa, and  
to Escherichia coli adhesin Aida-1 precursor SW:AIDA\_ECOLI  
(Q03155) (1286 aa) fasta scores: E(): 0, 29.2% id in 1286  
aa. Contains a possible N-terminal signal sequence."

/codon\_start=1

/transl\_table=11

/product="putative autotransporter protein"

/protein\_id="CAC89847.1"

/db\_xref="GI:15979073"

/translation="MNTIFKVIWNASLNVVSVSELAAGRIKTKSRNLISEGLPKF  
EQSMVSKLFRKLLASLGSIVFLSTGPFAADITVSTQAEASLANSNTGYDKIILGA  
DITLIGSLTVNMTSNQVWDGQKFGLTVNTNTYGLVSSGSGTILQNMKSIDSAN  
YYSVNLGANTAVNYNNIDFLGSSQLIYMGAYGAATNMTDFGDIYLVVNDRA  
QIEGVNKLATFREHYTHFGSSVTSFVSTGCCANNTSMDFASGADVKIDRGSTGDL  
TSTGVNAFYFADGASPELIANQVFSGTTNNGLEIGSYNSIDGFGSGVKVILQSLR  
SDGSIISGNDIATNAAGINNASGDANVYINLGTGSILKATNFIATLNANNAS  
DIYIRSGADITAAATGISTHNGTGVTKIKNDGTITSTAGIAISASIKESIVNTDG  
TITATAGTCVNVLASAILNFGGTTINTSATANGITFAGTEGGTITDILINLGTGIA  
LSNVAGVNLTSNVTLNLTNGTALNSLTGLTLVDSLNGRNTINIEGAGIATNTEL  
NTEFAEADINVGAGIGIATGCGGVNLSASNLIIINVANTLGLQITDGLIDNTTIG  
NEIQLNAENATINFLGSSKTLNNGTIKGSVIPAGVADHIIINNGTLDGLITTCAG  
NDTLVDSSSQSDNVINLGGNSVTIQNGATVSSITINGNDITFTINGMSGSTTLG  
SLDAGTGLNTLNASTDELAAATSLQGTNTINLVDSTHTLVSDNIGSGMWNIDSS  
ELLFGSTFDGLIHATLGAGTSATVNSANVLEQASMEFAGTVOVNOGQALFASNSQ  
LGSAKIGLDGLTLDNLIALFNHVLTCNGTLNVAKNLTAFTDFGSTVGAFSGVNLIT  
KTFALSADNAAALASATLKLSDDSVTVTGTDRTLHGLDLSGTLIFDGAVQSOTS  
GVYVTDLANSCTVNIIGSGSWDNTPLATNVSILEQDRAGSTLELINATNVTGDI  
ADLLVNGTATISGTQVQSAIQGGSTVANAIHNYGLAASNSDGLVNYLTSAL  
ELLADGALLLLEATSLTANRLNAELFVGGLVVDVDAONGALTLANGSNRYEGTTV  
TAGELILGANGAFGQTSLLIDTASGANINGYQTVGAVNTGVTGLSGSGVTLGSL  
TNGGILDLTGALNLTAGASTVAGGLTGAGTLNNGNLVSANSGSOGTHIADV  
ASVTLDTGTIGTSAVEVLTLNLANAATNVLSDGTINTNAVTLSGNSPSGA  
HOITDGLTVQASNLGASATVNLGTLTSHLILNGVSESANVLGSGVAGSTVDIG  
CADRALTANNSGFLQYALAGNSKLTAVSTNNLGGASSVALAGDGLSLSGFNFTIG  
NSVTSGVQLVTDDEAVTLTSSGVSNAVTIDATLNLDDILFNLHVLTNGLNLV  
AKNDASTDFGTSVGGAFSGIVNTTTFALSADNAAALATLKLSDDSVTGAT  
DRTLHGLDGLTILFDGSPQSQGVNVVTDLALNSGTISITGAGNENEPHTVPP  
NVSLLEQDRGDLLELINAANTVQANVVDLDTGTAITSGTGYESAQOQSGTVAN  
AIHNYGLTSSNGSGGLVYVTLTSALELLANGANALLATGSLTANRLNAELFVG  
YRQTVAVNSGAVTLGNGVGLTNGGLTGGALNDAAGSSVAVAGLTGAG  
TLNNGGDLAVSATNSLGGSTHIAVASVTLTGTGTGTSAYELVLTNGLNANAM  
TNVLSGGVINTNAVTLSSGNSFSAHQIGTDTGELTVQASNLGASSTVNLGTLT  
HLIINGSESTANVLGSGVAGTIIIGDADTALTANNGLFGYALAGSKLTGNTV  
NLGASSVALAGDGLSLSGFNFTGSGVGLVQVTDDEAVTLTSSGNSPSGA  
DIADATLNLDDIALFDHVLTCNGTLNVAKNLTAFTDFGSTVGAFSGVNLNTTFA  
LSADNAAALATLKLSDDSVTVTGTDRTLHGLDGLTIFDSSPQSQGVNVV  
TDLANSCTISITGAGNENEPHTVPPNVSLLEQDRGDLILQLIDANVTGNANDEL  
MINTTISAGQVQSTVQGGVYVANATHYNGMTSGGGLVYNTLSALELLADGAN  
ALLATGSLTANRLNAELFVGGLVDAONGALTLANGNRYEGTTVTAAGELTLG  
ANGAFQTSLLNLAGSASANTYRQTVAVNTGVTGTGNGELTSTDTLNTGMIN  
VTDLNLNENGASISGLTGNGILNKGCDFTISIDNGLAGOTNIDSGASTLGN  
GGTILGTNLGSSVIDVLCDLNLVADNLSANVISDGTINTATVTLSGNSFSGAHQ  
IGTNGELTVQASNLGASSTVNLGTLTSHLILNGVSESANVLGSGVAGSTVDIGGA  
DTALTANNSGFLQYALAGNSKLTAVSTNNLGGASSVALAGDGLSLSGFNFTGNS  
NDASTDFGATVCGAFTGVNLTNNSTFDLSGNNTTTLAQATLKLSSGNLTVSGVNGV  
NIGTLAMNGTLLFDNIYDAGITSDGTIANSINTTGGGEVRYNLPSNLAPSLDGL  
SVNELDEGEIIVLTATGAATGTGHELTIDNDGPIISAVTYQGVNDSIAAQTSAGTGF  
NYGTMGTGDEGLVYNYVNTALELSTGSEALVLTALANNGTQNDLSAQTSAGTGSGL  
AFASDGTASLNSNTSYGTWVSSGNRLDADSLAGQTSLLAMSTAYDNDAGTGTG  
QOVVGEATGEGSTLDLNDGKLTVTGGQIDGALGGGELVGLGSLNANVYDNGAFTG  
STDANGVAHLSQAQGLNGCTINNGTLHLDNTGTLFNALTGSDGEVLLSNNSAVQ  
LAGDNGSGYGLFTNQAGSLIANSAGHLGGSSIANSGALIDTSGWEIWTNTISGTGT  
LVKRGSGTGTGEGDTSAGLTIEGLLIQAGSSAVTQTLSEESLQERALSFSASNM  
ANLTSNLTANGSLGGYGTQVAVNTYNNIIMPALTGDFGTIDGNTGTDEGMI  
TFNTILAGDTSVTDRLVTGTAGQSVYVNNIGVGARTFEGIKIIDVGGDSAQGT  
LNGRAGVGAEVFLYOGCASTPDDGNVYLTEDDRPERPASYTANLAANNMFVLSL  
ADRNGETLYTVDFGQKTLTSLWLNESGHNSRSDSGELKTDQNRVYVMOGGDVAQW  
SRNAQDLNRVGVGMAGYNSSTVAQVAGYSTGVSDGYSVGIYGSWLNADDDTGYAY  
VDSWQVSWFNDRVSGDQKATEKYDSKGTASVEGGYAFKVESYQSYFTQPKAQVY  
WMGVKADDDHTNCTVISGDGNGNIQTRLGAKAFINPDKARVSGPAPKPFVEANNIH

```

NTKDFGTTLDGVTVKQAGTANIAELKLGVDGQVNSQLNLWNLGNIGQQVGNKGYSVTV
LGKYNF"
/note="complement(12049. .13929)
/gene="yp01005"
CDS
complement(12049. .13929)
/gene="yp01005"
/note="Similar to neighbouring CDSs yp01007 and yp01006.
Similar to Salmonella typhimurium leucine-rich repeat
protein SLP TR:O9XCV2 (EMBL:AF127079) (765 aa) fasta
scores: E(): 0. 39.0% id in 644 aa, and to Shigella
flexneri 65.4 kDa antigen IpaH 4.5 SW:IPA4_SHIPL (F18009)
(574 aa) fasta scores: E(): 0. 40.4% id in 589 aa"
/codon_start=1
/transl_table=11
/product="putative antigenic leucine-rich repeat protein"
/protein_id="CAC89848.1"
/db_xref="GI:15979074"
/translacion="MNLSTNSVSNMNPIDREIHSARTSTAALTPADYVIAIWEKWE
SEAIPTGDEQRRAVECMKDCLEKNTYHLCLRDNLASLPDILPPCNELDMGNKLTVE
LPATLPDNLQKLNASFQNLRTLPDTPSLSLSNVYGNELERLPESLPEGLKELDVND
NESLQLPNLPNLPNLPNLPNLPNLPNLPNLPNLPNLPNLPNLPNLPNLPNLPNLPNLPN
SVTNSQNLQTPNLPNLPNLPNLPNLPNLPNLPNLPNLPNLPNLPNLPNLPNLPNLPNLPN
PQCEILLEGPNLSTSLQVQLHLRINPYQGRINWSELDNPPASLRNIVATWLP
EQNRLAGDANITEANSAAFSVFLHRLATTQNNANPEFKQOIAAWLQLADSPFL
REQFLTAQASATCEDRITLTHNDQKAVMLHEVEKGYDEKLPMLARGEMFRLE
QLENIAREKVIILKTLNVNSVDIEVYLAYQVLLNSLQSLSSVKNEMRFFGVSHVTAD
DLLSAETRVKTAENQDPSRWLSQSPKSVVQRIEPRYAAAEKQVHALENTYPPDKL
AAELAAANGMGDGDANRIVKGRINDELGMELDMALTHEVLSKAGSSLLDNLWMEYLI
Sp"
complement(13099. .13164)
/gene="yp01005"
/misc_feature
/note="Pfam match to entry PF00560 LRR, Leucine Rich
Repeat, score 9.30, E-value 44"
complement(13219. .13290)
/gene="yp01005"
/misc_feature
/note="Pfam match to entry PF00560 LRR, Leucine Rich
Repeat, score 12.80, E-value 8.4"
complement(13294. .13353)
/gene="yp01005"
/misc_feature
/note="Pfam match to entry PF00560 LRR, Leucine Rich
Repeat, score 10.40, E-value 30"
complement(13468. .13539)
/gene="yp01005"
/misc_feature
/note="Pfam match to entry PF00560 LRR, Leucine Rich
Repeat, score 17.30, E-value 0.36"
complement(13543. .13602)
/gene="yp01005"
/misc_feature
/note="Pfam match to entry PF00560 LRR, Leucine Rich
Repeat, score 12.90, E-value 7.8"
complement(14492. .15367)
/gene="yp01006"
gene
/note="Similar to neighbouring CDSs yp01005 and yp01007.
Similar to regions of Salmonella typhimurium secreted
protein h2 TR:Q9RPH0 (EMBL:AF160727) (788 aa) fasta
scores: E(): 2.5e-24, 41.9% id in 272 aa, and to regions
of Yersinia pestis outer membrane protein YopM TR:O68701
(EMBL:AF053946) (409 aa) fasta scores: E(): 7e-23, 40.5%
id in 259 aa"
/codon_start=1
/transl_table=11
/product="putative antigenic leucine-rich repeat protein"
/protein_id="CAC89849.1"
/db_xref="GI:15979075"
/translacion="MYLSTNSVSNMNPIDREIHADRPAAATLPADYVIAIWEKWE
NDPRTVAGRGQAVAKMCELNNAERLNLSSDLTSLDTPLPNCNLICNNLTE
LPTLPDNLQKLNASFQNLRTLPDTPSLSLSNVYGNELERLPESLPEGLKELDVNC
NTSLQPSRLPPVLESIDNSICNLTLPNLPNLPNLPNLPNLPNLPNLPNLPNLPNLPN
SVAYNOITLPNLPNLPNLPNLPNLPNLPNLPNLPNLPNLPNLPNLPNLPNLPNLPNLPN
LRLSTKPNYQGRIS"
complement(14657. .14728)
/gene="yp01006"
/misc_feature

```

```

/note="Pfam match to entry PF00560 LRR, Leucine Rich
Repeat, score 15.40, E-value 1.3"
complement(14732. .14791)
/gene="yp01006"
/misc_feature
/note="Pfam match to entry PF00560 LRR, Leucine Rich
Repeat, score 9.30, E-value 43"
complement(14981. .15040)
/gene="yp01006"
/note="Pfam match to entry PF00560 LRR, Leucine Rich
Repeat, score 11.30, E-value 22"
alignment_scores:
Quality: 1265.50 Length: 826
Ratio: 2.256 Gaps: 14
Percent Similarity: 67.918 Percent Identity: 35.472
alignment_block:
US-09-701-711-2 x AJ414146
Align seg 1/1 to: AJ414146 from: 1 to: 210050
10 GlnValSerAlaMetThrMetAlaValMetMetValMetSer..... 23
74360 CGCATAACAACGATGGCGATGAAAGTTGCTCATAGCGTCGCTGCTGTT 74409
24 .....ThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIleT 37
74410 TGGCAGCGCCACCGTATACGGTGCAGCGGGTTCGTAGTGAACGATATTC 74459
37 hrIleThrGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeu 53
74460 ATTTCGAAGGACTGCAACGGGTGCGCGGTGCTTACTTAATATG 74509
54 ProPheArgLeuGlyValValSerGluAsnGlnLeuAlaAspGlyVa 70
74510 CCGGTTCGGTAGGCGATACCGTCAGTGATGATGATCGGTAAACTAT 74559
70 LysAlaLeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisG 87
74560 CCGTTCGGTGTGTTGCTACAGCAACTTCGAGGACGTCGCGCTCGCGG 74609
87 InGluGlyArgIleIleTyrGlnValThrGluArgProLeuIleAlaGlu 103
74610 ATGGTAATACGCTGATTGTTCAAGTCAAAGAACGCCACGATGCGCAG 74659
104 IleAsnPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluG 120
74660 ATCACTTTCGGGTAAATAACGGCGTGAAGAAGATATGCTCAACAGAA 74709
120 YLeuLysAsnAlaGlyLeuAlaValGlyGlnProLysGlnAlaThrV 137
74710 TCTGGAAGCCTCTGCGTCCGGTGGCGGCTTGGACCGAACCACTA 74759
137 alGlnMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyr 153
74760 TATCAACATCTGAAAGAGCAGCTCGAAGATTTCTATTACAGTGTGGGTAA 74809
154 TyrAsnThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgVa 170
74810 TACACGGCTTCAGGCTGTTGTTAGCCATTACCACGTAATCGTGT 74859
170 llysLeuAspMetThrPheAlaGluGlyLysProAlaArgValAspI 187
74860 TGACCTGAATTTGGCTTTTACCGAAGGTGTTTCTGCAAAATTCAGCAGA 74909
187 leAsnIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspVal 203
74910 TTAATATCGTAGTATCACAGTTTACCACCTGATGAGCTTATCTCACGC 74959
204 LeuAlaIleLysAsp.....AsnLysIleAsnProLeuSe 215
74960 TTCCAACCTCGGTGACGAGGTGCTTGGTGGAAACGATAGTC..... 74998

```



```

76729 ATTCAGTTTAAACATTGGTAAACTTGG 76756

seq_name: gb_ba:PLU236920

seq_documentation_block:
LOCUS      PLU236920               4405 bp    DNA         linear   BCT 02-AUG-1999
DEFINITION Photorhabdus luminescens yael (partial), fira (partial), oma and
              ompH genes.
ACCESSION  AJ236920
VERSION    AJ236920.1   GI:5689864
KEYWORDS   fira gene; oma gene; ompH gene; outer membrane antigen; periplasmic
              protein; yael gene.
SOURCE     Photorhabdus luminescens.
ORGANISM   Photorhabdus luminescens.
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Photorhabdus.
REFERENCE  1 (bases 1 to 4405)
            Chatonnet-Marton,P.I., Givaudan,A., Lanois,A. and Boemare,N.E.
            Photorhabdus luminescens genomic region homologous to 4.0 minute
            Escherichia coli region promotes pleiotropic phenotypes
            Unpublished
            Direct Submission
            Chatonnet-Marton,P.I.
            Submitted (12-FEB-1999) Chatonnet-Marton P.I., Laboratoire de
            Pathologie Comparee, Institut National de la Recherche Agronomique,
            Universite Montpellier II, C.C.101, Pl. E. Bataillon, Montpellier,
            34095, FRANCE
FEATURES             Location/Qualifiers
     source           1..4405
                     /organism="Photorhabdus luminescens"
                     /strain="Hm"
                     /db_xref="taxon:29488"
     gene             1..681
                     /gene="yael"
     CDS              1..881
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB51928.1"
                     /db_xref="GI:5689865"
                     /translation="MEKVIPGSAEAKAGLQKDRIVKVGQSEIDVWHITFTSFVSNPNP
                     VLEUVRAGHIISLSTMPVROGGRKVGAGVELRIIVPLADETKIVQKGPESA
                     MYQADKTVTQMLKRLTIVGMIGLVGDVKINNLSPISIAKGAGVSDSGSLVLYLMFLA
                     LISVNLGIINLPLVLDGHLLELFTIEKINGPVSERVQDFSYRIGAMILVLLMGLA
                     LFNDFSRF"
     RBS              696..703
                     /gene="oma"
     gene             join(696..703,716..3109)
                     /gene="oma"
                     /gene="oma"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="outer membrane antigen"
                     /protein_id="CAB51929.1"
                     /db_xref="GI:5689866"
     CDS              716..3109
                     /gene="oma"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="periplasmic protein"
                     /protein_id="CAB51930.1"
                     /db_xref="GI:5689867"
                     /translation="MKKILCAASFGIALAFSAGAQAADKIAVVNVGGEIFQQLPAREAV
                     AKLENEFKNRASLQRMETDLSKIOKLRDSTWKKSSERTNLEKEVMAKKEEFKAK
                     AQAEQDHRREMEERNKILSIQDAIKVVAGREGDIVIDANAVAIYSVGNKITASY
                     LKQVK"
     gene             3703..3709
                     /gene="fira"
     RBS              3703..3709
                     /gene="fira"
     gene             3721..4405
                     /gene="fira"
     CDS              3721..4405
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="UDP-3-O-(3-hydroxymyristoyl) glucosamine
                     N-acetyltransferase"
                     /protein_id="CAB51931.1"
                     /db_xref="GI:5689868"
                     /translation="MSSIRLADLAQQLNAQLHGDGDIVITSIAPMHSANSEQITFLSD
                     SYRERLGECAAAVLAQASDLPCYNIPALVWNPYLAAYMAQIMDTTPIPAQDIHL
                     SAVLSPOATGKNVAVGANAVIESVVLGDNVVGACGFIGKNTKTRIGAGSRLWVWVSV
                     YHNVMGEQCLIQSGTVIGSDGFGYANDRGKKWKIPOLSSVIMGDRVEVGACTTIDIR
                     TLDNRIIIGNG"
BASE COUNT      1187 a   819 c   1094 g   1305 t
ORIGIN
alignment_scores:
    Quality: 1258.50      Length: 826
    Ratio: 2.223          Gaps: 13
    Percent Similarity: 68.523    Percent Identity: 34.363
alignment_block:
US-09-701-711-2 x PLU236920 ..
Align seg 1/1 to: PLU236920 from: 1 to: 4405
8  GlypHeGlnValSerAlaMetThrMetAlaValMetMetValMet..... 22
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
698  GGAAGAGCGCATTAAACAGTGGCGATGATAAGTAAAGTTACTCATGCGTCGCT 747
23  .....SerThrHisIaGlnAlaAaasp....PheMetAlaAsnA 35
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
748  GCTGTTGCGGAGCGCGCTGCATACCGTGCAGCGGATTCGTAGTTCAAG 797
35  splitHeThrIleThrLeuGlnArgValThrIleGluSerLeuGlnSer 51
   ||||| : : : : : ||||| : : : : : ||||| : : : : :
798  ACATTCACTTTGAAGTCTTCACGCGTCGCCGTAGTGCAGCAATTATTTG 847
52  ValLeuProPheArgLeuGlyGlnValValSerGluAsnGlnLeuAlaAs 68
   : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
848  AATATCGCACTTCGCGTAGTGATACGTCAGCGATGAAGATATCGTCG 897

```

365 sThrGlnAspGluValLeuArgArgGluMetArgGlnLeuGluGlyAlaAl 382  
 1792 CACTAAAGATCCCGTCTCGCGTCGTGAATCGCTACAGTGAAGAAGACAT 1841  
 382 euAlaSerAsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArgThr 398  
 1842 GGTTAGGAGATGATTAGTGTGAACTAGTGAAGAAGCGCTTTACCGTTG 1891  
 399 GlyPhePheLysHisValThrValAspThrArgProValProAsnSerPr 415  
 1892 GCTTATTTCGAACCGTAGATGTTCAAACTCAGCGTATACCGCAGCC 1941  
 415 oAspGlnValAspValAsnPheValValGluGluGlnProSerGlySerS 432  
 1942 TCGTCAGGTCCGATCGTGTATTAAGAGTTAAAGAACGTAATACCGTCTC 1991  
 432 erThiLeuAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPhe 448  
 1992 TGAACCTTTGGGTCTGGTTCGTACAGAAGCGCGTTAGTTTCCAAT 2041  
 449 AspValSerGlnAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSe 465  
 2042 GCGCGTCAACAGGATACTGGTGGGTACAGGTAAACCGCTGTGGAATTA 2091  
 465 rPheSerArgSerGluThrArgGluValTyrSerLeuGlyMetThrAsnP 482  
 2092 TGCCAGTAAAGACGATTTATCAACTTACGCAAGCTCTCCTTCACATG 2141  
 482 roTyrPheThrValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArg 498  
 2142 CATACTTTACGATTAAATGGTGTGAGCCCTGGTGGTGCAGATTTCTACA 2191  
 499 LysThrLysTyrAspAsnLysAsnIleSerAsnTyrValLeuAspSerTy 515  
 2192 GATTTTCAGACCGCATGATCGTAATATTCGGTTACACTTAACCAAGCTA 2241  
 515 rGlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleS 532  
 2242 TGGTATAAGCGGTTTCCCTTGGTTTTCCCGATTAAATGAATAAATACTCG 2291  
 532 erPheGlyLeuAsnAlaAspAsnThrLysLeuHisGlyGlyArgPheMet 548  
 2292 ATTTTGGCTTAAATATATTCATAACTCTGTG...TCTGATATGCTTCCT 2338  
 549 GlyIleSerAsnValLysGlnLeuMetAlaAspGlyGlyLysIleGlnVa 565  
 2339 CAGGTAGCAATGTGCGGTATCTGAGGTCTATGGCCGACGAACCTGATCT 2388  
 565 lAspAsnAsnGlyIleProAspPheLysHisAspTyrThrTyrAsnA 582  
 2389 TGAAGATTAAGCC...GAATTCAAAGCAGAT...GATTTTGCTC 2426  
 582 laIleLeuGlyTyrAsnTyrSerSerLeuAspArgProValPheProThr 598  
 2427 TCACATATGGATGGACATATACCAACCTCGACCGTGTCTTCTCCCAACA 2476  
 599 GlnGlyMetSerHisSerValAsp...LeuThrValGlyPheGlyLys 613  
 2477 TCAGGTGGAAGTCGACTCTGAATGGTAAAGTGAACGTACTTCTGGTCCGA 2526  
 613 pLysThrHisGlnLysValValTyrGlnGlyAsnIleTyrArgProPhe. 629  
 2527 TAACGAATTTTACAAAGTGAATTTGGAGCTAGCCGTTTAGGGTATGGG 2626  
 645 Asn.....AsnLeuProPheTyrGluAsnPheTyrAlaG1 656  
 2627 GATGATTAGCGCGTGAAGAGTGTCCCAATTCATGAAGAACTTCATGCAGG 2676  
 656 yGlyTyrGlySerValArgGlyTyrAspGlnSerSerLeuGlyProArgS 673



predicted by Codon\_usage  
 predicted by Homology  
 predicted by Framed"

/codon\_start=1  
 /transl\_table=11  
 /evidence=not\_experimental  
 /product="PROBABLE THREONINE SYNTHASE PROTEIN"

/protein\_id="CAD15030.1"  
 /db\_xref="GI:17428343"

/translation="MQYRSTRGHAEPQSFSSRILLGLADPDGGILPEQYPOVSAEELT  
 RWDLPLADLAEILKRFATDIPEDDLRALTRRTYPEVYCARNQNDATITPLHRL  
 EGAGTSLSLNSGNTLAFKDMAMQLLGNLFYALDREHAELNLGATSGDTGSAA  
 EYAMKRGVRFVFMELSPHRKMAFQTAQMFSLQDKNFNIAIEGVFDDAQDIKAVSN  
 DLEYKARQIGTVNSINARVVAOVVYFKGYLLAAPKIGDKVFCVPSGNGVNCAG  
 HTARMGLPIDKLIVATNENDVLDEFRTYVRSAOQYHTSPSMDISKASNEER  
 VFYDILGRPARLAQLFRVDYDEKGFSGMTGKPEFDRIAEFGVSGRSTHEDRVNTRD  
 VFYRYGTDVTHGKVAKENLVPGVMVLETPALPAKFGDTIHEALDREPERPAG  
 YADIESLPQRFVEYMPADADRIRKGYIAQHTGL"  
 4361..4957

gene

/gene="RSC1329"

/note="RS02851"

4361..4957

/gene="RSC1329"

/function="miscellaneous; hypothetical/global homology"  
 /note="Product confidence : hypothetical

Gene name confidence : hypothetical

predicted by Codon\_usage

predicted by Homology

predicted by Framed"

/codon\_start=1

/transl\_table=11

/evidence=not\_experimental

/product="CONSERVED HYPOTHETICAL PROTEIN"

/protein\_id="CAD15031.1"

/db\_xref="GI:17428344"

/translation="MAEAFASCLHDIRACRICEGALPHEPRPVVLADRRARVLIVGQ  
 AGRIVHETGIPWNRSGDRLEWLRIGYDAFYADRIAVVPMFCFPGTGNGLDLP  
 RPACAPQWRPLLTAMPDLKLALLGVGYAQRYVLPVPRGMPLTEVVRGPFVDAEGRV  
 LPLPLPSPRNLAWEKHPWFDAQIVPLRLAALAAAFV"  
 5037..6296

gene

/gene="moa2"

/note="RSC1330; RS02852"

5037..6296

/gene="moa2"

/function="small molecule metabolism; biosynthesis of  
 cofactors, carriers; molybdopterin"

/note="Product confidence : putative

Gene name confidence : putative

predicted by Codon\_usage

predicted by Homology

predicted by Framed"

/codon\_start=1

/transl\_table=11

/evidence=not\_experimental

/product="PUTATIVE MOLYBDOPTERIN BIOSYNTHESIS MOEA  
 PROTEIN"

/protein\_id="CAD15032.1"

/db\_xref="GI:17428345"

/translation="MSANDAPAKVPSLLTLEALLLOLLAAARGLTETVPTLHANGR  
 VLAQAVRSGLNPVADNDQMDGYAVRAADITAPGTLKYQRIIPAGVGRPLQAGEAA  
 RIFTGALIPPGADAVVMEQCKPDDPTGTVDHVPKEAGEWIRRRAGEDITAGAEILPR  
 GTRLGAQGLLAASVGCAGLVQVRRPRAVFFTGDELAMPGEPLKPGAIYNSNRFLLR  
 ALLENIGCEITDFGIVPDTLOARTDLREAAEADHDIITSGGVSVGEEDHIKPAVEAE  
 GRNLWQIAIKRIPKPLAFGEVRRPQGMAGGPAFFTGLPGNPVSSFTVLLFVRPFI  
 RLGLVDVTPKRIPMRADFALPKGRNEFLRINAGGGLDLFPNQSSGLTSTVWG  
 DGLINDPNQPTAGDTVAFLPFAGLV"  
 6310..6573

gene

/gene="moa2"

/note="RSC1331; RS02853"

6310..6573

/gene="moa2"

/function="small molecule metabolism; biosynthesis of  
 cofactors, carriers; molybdopterin"

/note="Product confidence : probable  
 Gene name confidence : probable  
 predicted by Codon\_usage  
 predicted by Homology  
 predicted by Framed"

/codon\_start=1

/transl\_table=11

/evidence=not\_experimental

/product="PROBABLE MOLYBDOPTERIN MPT CONVERTING FACTOR  
 (SUBUNIT 1) PROTEIN"

/protein\_id="CAD15033.1"

/db\_xref="GI:17428346"

/translation="MDMQIELRFFASVREOLGTSHEAAAAPDVTVRTVGEILRLWLAARG  
 PAWGETLAEGRALRMALDHDVVVTADAPLREGCEVAFPPVPTGG"  
 6586..7116

gene

/gene="moaE"

/note="RSC1332; RS02854"

6586..7116

/gene="moaE"

/function="small molecule metabolism; biosynthesis of  
 cofactors, carriers; molybdopterin"

/note="Product confidence : probable

Gene name confidence : probable

alignment\_scores:

Quality: 1237.00 Length: 823

Ratio: 2.257 Gaps: 17

Percent Similarity: 66.586 Percent Identity: 34.022

alignment\_block:

US-09-701-711-2 x AL646064 ..

Align seg 1/1 to: AL646064 from: 1 to: 201050

9 PheGlnValSerAlaMetThrMetAlaValMetMetValMetSerThrHi 25

100731 TTCCCGCTCAGCATGCTGGCGGCTTCCGTGTGACCGTTGCGCGGTCA 100780

25 sAlaGlnAlaAlaAsp...PheMetAlaAsnAspIleThrIleThrGlyL 41

100781 GGCTCATGTCAGTCGAGCGGCTTGGTCATCAAGGACATTCGCTGGAGGGGG 100830

41 euGlnArgValThrIleGluSerLeuGlnSerValLeuPropPheArgLeu 57

100831 TGCAACGCGTCGAGCGCGCACCGCTATTTCGGCTATCTCCCGTGAAGGTG 100880

58 GlyGlnValValSerGluAsnGlnLeuAlaAspGlyValLysAlaLeuTy 74

100881 GGTGAGACCTTCACCGAGCAGACAGGCGCGCAATCGATCGCGCGCTGTA 100930

74 rAlaThrGlyAsnPheSerAspValGlnValTyHisGlnGluGlyArgI 91

100931 CAACACCGGTTTCTCAAGGATGTACAGATCGCGCGGAGGCAACGTGC 100980

91 leIleTyThrGlnValThrGluArgProLeuIleAlaGluIleAsnPheGlu 107

100981 TGGTGGTCCGCGTGGAGGAGCGTCCGGCAATCTCCACAGCTGGAATTCATC 101030

108 GlyAsnArgLeuIleProLysGluGlyLeuGlnGluGlyLeuLysAsnAl 124

101031 GGCAATCAAGGAATTCGACACGACACGACGCTGCGCGGCTGCGCGGCGG 101080

124 aGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrValGlnMetIleG 141

101081 GGGCGTGCCCGAGCGCGCTACTACGACAATCCCTGATCGACCGTGGCG 101130

141 luThrGluLeuThrAsnGlnTyTrIleSerGlnGlyTyTrAsnThrGlu 157

101131 AGCAGGAGCTCAAGCGCCAGTACGTCTCGCGCGGTACTACGCCCGCCGAC 101180

158 IleThrValLysGlnThrMetLeuAspGlyAsnArgValLysLeuAspMe 174

101181 GTGCAGACCCACCGTCAGCGCCAGTCTGATGCCAACCGGGTGTCCGTGTCACGT 101230









```

/misc_feature
/transl_table=11
/product="ribosome recycling factor"
/protein_id="CAD08677.1"
/db_xref="GI:16501500"
/translacion="MISDIRKDAEVRMEKVEAFKIQISKVTRGRASPSLLDGIVVEY
YGTPTPLRQLASVTIVDSRLTKINVDKSMGPAVERKAIMASDLGLNPSAGTDIRVPL
PPLTEERKDLTKIVRGEAQAVRVRNRDANKVKALLKDKAKAISEDDEDRRSQEEV
OKMTDAIKKVDRAADLADKEALMQF"
3086..3580
/gene="STV0242"
/notes="Pfam match to entry PF01765 RRF, Ribosome recycling
factor, score 339.90, E-value 2.9e-98"
3731..4927
/gene="dxr"
/notes="STV0243; yaem"
3731..4927
/gene="dxr"
/EC_number="1.1.1.-"
/notes="Similar to Escherichia coli 1-deoxy-D-xylulose
5-phosphate reductoisomerase dxr SW.DXR_ECOLI (P45568;
P77209) (398 aa) fasta scores: E(): 0, 88.9% id in 397 aa
Orthologue of E. coli yaem (DXR_ECOLI); Fasta hit to
DXR_ECOLI (398 aa), 89% identity in 397 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein in frr 3' region"
/protein_id="CAD08678.1"
/db_xref="GI:16501501"
/translacion="MKQLTILGSTGICSTLDVYHNPDSFRVIALVAGKNVMAE
QCLEFSRYAVMDTSSAQKIMLQOHGSRTEVLGQQAACEMAAALDEGVHMAAIV
GAAGLLPTAAIRAGKITLLANKESLVTGRLFMDEKRSNARLLPVDSEHNAIFOSL
POSTQHNLTGYADLEQNGVTSILLTGSGGPERETPCMDLAAMTPDOACRHPNWSMGRI
SVSATMNGKLEXYEARWLFNASARQMEVLIHPQSVIHSWRYQDSVLAQLGEPM
RTPAHMAMPNRTSGAQDLDFCKLSALTFSPADYQRYFCLKLAMEAFEGQQAQYTA
LNANETIVAAFLAQOIRFTDIDAGLNLAFLVRLMDLHEPASVDDVLDVQDATAREVARQ
VIRLSR"
4064..4087
/gene="dxr"
/notes="PS00017 ATP/GTP-binding site motif A (P-loop)"
5240..5998
/gene="STV0244"
5240..5998
/gene="STV0244"
/notes="Orthologue of E. coli UPPS_ECOLI; Fasta hit to
UPPS_ECOLI (253 aa), 94% identity in 252 aa overlap"
/codon_start=1
/transl_table=11
/product="undecaprenyl pyrophosphate synthetase"
/protein_id="CAD08679.1"
/db_xref="GI:16501502"
/translacion="MLSATQPVSNLPAHGCRHVAIIMDGNRWAKQKIRAFGHKA
GAKSVRRVAFNANGIDALTLYAFSSNNWRNPAQEVSAIMELFVWALDSEVKSLSRH
NVRRIIGDTSRNSRLQERIRKSEALTHTNTGLTINIAANYGRWRDIVOGVRLAEL
VQAGVLRPDQIDERLQOQCMHELAPVDLIVRTGGEHRISNELLWQIAELYFTDV
LWPDFDQDFEGALHAFNRFRPGGTEPGDDKA"
5306..5971
/gene="STV0244"
/notes="Pfam match to entry PF01255 UPP_synthetase,
Putative undecaprenyl diphosphate synthase, score 492.70,
E-value 2.8e-144"
5804..5857
/gene="STV0244"
/notes="PS01066 Undecaprenyl pyrophosphate synthetase
family signature"
6011..6868
/gene="STV0245"
/notes="cdsa"
6011..6868
/gene="STV0245"
/notes="Fasta hit to YNBB_ECOLI (298 aa), 33% identity in
298 aa overlap
Orthologue of E. coli cdsA (CDSA_ECOLI); Fasta hit to

```

```

alignment_scores:
  Quality: 1220.50      Length: 845
  Ratio: 2.211         Gaps: 15
  Percent Similarity: 65.335 Percent Identity: 33.728

alignment_block:
US-09-701-711-2 x AL627266 ..

Align seg 1/1 to: AL627266 from: 1 to: 268050

10 GlnValSerAlaMetThrMetAlaValMetMetValMetSer..... 23
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
8252 CGCATAATAACGATGCGCATGAAAAAGTTGCTCATAGCGTCGCTGTT 8301
24 .....ThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIleT 37
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8302 CAGCAGCGCCACCGTATACGGTGTGAAGGGTTCGTAGTGAAGGACATTC 8351
37 hrIleThrGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeu 53
   ::::: ||||| ||||| ||||| ||||| ::::: ::::: :::::
8352 ATTTCGAAGGCTCGCAGCGAGTCGCGGTGGTGGCGCTCTCCAGTAGTG 8401
54 ProPheArgLeuGlyGlnValValSerGluAsnGlnLeuAlaAspGlyVa 70
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8402 CCGGTGCGCACAGCGCACACGGTTAATGATGAAGATATCAGTAACACCAT 8451
70 llyAlaLeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisG 87
   ::::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
8452 TCGCGCTCTGTTGCGCACCGCACTTTGAGGACGTCGCGCTCCTCGCGG 8501
87 InGluGlyArgIleIleTyrGlnValThrGluArgProLeuIleAlaGlu 103
   ::::: ::::: ||||| ||||| ||||| ||||| ||||| |||||
8502 ATGGTAATACCTTCTGGTTCAGTAAAGAACGTCGACCATTCGCCAGC 8551
104 IleAsnPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluGl 120
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8552 ATCACTTTCTCGGGAACAAGTCGGTGAAGATGACATGCTCAAGCAAAA 8601
120 vLeuLysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrV 137
   ||||| ::::: ||||| ||||| ||||| ||||| ||||| |||||
8602 CCTCGAAGCGTCTGGGTACGTTGGCGAGTCTCTGGATCGCACACGC 8651
137 alGlnMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyr 153
   ::::: ||||| ||| ::::: ||| ||| ||| ||| ||| |||
8652 TGTCTGATATCGAAAAAGTCTCGAAGACTTTTACTATAGCGTAGGTAAA 8701
154 TyrAsnThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgVa 170
   ||||| ::::: ||||| ||||| ||||| ||||| ||||| |||||
8702 TACAGCGCAGGTTAAGCGGTGCTCAGCGCGCTACCGCGCAACCGTGT 8751
170 llyLysLeuAspMetThrPheAlaGluGlyLysProAlaArgValAlaSpI 187
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8752 TCATCTAAACTGGTGTTCAGGAGGCGGTATCGGCGAAGATCCAACAGA 8801
187 leAsnIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspVal 203
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8802 TCAACATCGTCGTAACTGCTTTAGCACCGCAAGAACTCATTTCTCAT 8851
204 LeuAlaIleLysAsp.....AsnLysIleAsnProLeuSe 215
   ::::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
8852 TTCCAGTTCGTGACGAAGTGGCGTGGTGAACGTCGTC..... 8890
215 rLysAlaAspArg...TyrThrGlnGluLysLeuValThrSerLeuGluA 231
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8891 .....GGCGATCGTAATACAGAAACAGAAAGCTGGCGGCGACCTTGA 8936
231 snLeuArgAlaLysTyrLeuAsnAlaGlyPheValArgPheGluIleLys 247
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8937 CCTTGGCGAGCTACTATCTGGATCGCGGCTACCGCGCTTTCAATATTG 8986
248 AspAlaLysLeuAsnIleAsnGluAspLysAsnArgIlePheValGluI 264
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8987 TCCACCAAGTGAGTCTGACGCGGATAAGAAAGGATCTACATCACCGT 9036

```

264 eSerLeuHisGluGluGluGlnTyrArgPheGlyGlnThrGlnPheLeuG 281  
 : : : : : ||||| : : : : : : : : : : : : : : : : : : : :  
 9037 CAATATCACCAAGGCGATCAGTACAAGCTTCCGGGGTTCCAGTGAGCG 9086  
 281 lysAsnLeuThrTyrThrGlnAlaGluLeuGluAlaLeuLeuLysPheLys 297  
 : : : : : ||||| : : : : : : : : : : : : : : : : : : : :  
 9087 GTAACTGGCCGACATCTCCGCTGAAATTGAGAACTGACGAAATGAA 9136  
 298 AlaGluGluGlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsnI 314  
 ||| : : : : : : : : : : : : : : : : : : : : : : : :  
 9137 CCGGGCGAGCTCTATACGGCCACCAAGTACCAAAATGGAAGATGACAT 9186  
 314 eSerThrLysPheGlyAspAspGlyTyrTyrTyrAlaGlnIleArgProV 331  
 : : : : : ||||| : : : : : : : : : : : : : : : : : : : :  
 9187 CAAAAGCTTCTGGTGGCTATGTTACGCTTCCGCGCGTGCAGTCAC 9236  
 331 alThrArgIleAsnAspGluSerArgThrValAspValGluTyrTyrIle 347  
 : : : : : ||||| : : : : : : : : : : : : : : : : : : : :  
 9237 AGCTTGAATTAACGATGCTGCAAAACCGTAAACCTGCGCGTCAACGTC 9286  
 348 AspProValHisProValTyrValArgArgIleAsnPheThrGlyAsnPh 364  
 ||| : : : : : ||||| : : : : : : : : : : : : : : : : : :  
 9287 GATCGGGCAACCGTTTCTACGTCGCTAAGATCCGCTTGAAGGCACGA 9336  
 364 eLysThrGlnAspGluValLeuArgArgGluMetArgGlnLeuGluGlyA 381  
 : : : : : ||||| : : : : : : : : : : : : : : : : : : : :  
 9337 TACCTCAAAAGATTCCGTTCTCGCGCGTGAAATCGCCAGATGGAAGGG 9386  
 381 laLeuAlaSerAsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArg 397  
 || : : : : : : : : : : : : : : : : : : : : : : : :  
 9387 CGTGGTGGCGACGACCTGTGTGACCGAGGGGAAAGAGCGTCTGAACCG 9436  
 398 ThrGlyPhePheLysHisValThrValAspThrArgProValProAsnSe 414  
 ||||| : : : : : ||| : : : : : ||||| : : : : : : : : : :  
 9437 CTGGCTTCTTTGAACCGTCGACCGACCGACCGCGGCTTCCGGGAG 9486  
 414 rProAspGlnValAspValAsnPheValValGluGlnGlnProSerGlyS 431  
 ||||| : : : : : : : : : : : : : : : : : : : : : : : :  
 9487 CCGGATCAGGTTGATGTGGTGACAGAGGTGAAAGCGGTAAACACCGTA 9536  
 431 eSerThrIleAlaAlaGlyTyrSerGlnSerGlyGlyValThrPheGln 447  
 || : : : : : ||||| : : : : : : : : : : : : : : : : : :  
 9537 GCTCAACTTCGGTATCGGTACGGTACAGAAAGCGCGGTGAGTCCAC 9586  
 448 PheAspValSerGlnAsnAsnPheMetGlyThrGlyLysHisValAsnAl 464  
 : : : : : ||||| : : : : : : : : : : : : : : : : : : : :  
 9587 GCCGGCTTCAGCAGACAACTGGTTAGGTACCGGTTATCCCGTGGGAT 9636  
 464 aSerPheSerArgSerGluThrArgGluValTyrSerLeuGlyMetThrA 481  
 : : : : : : : : : : : : : : : : : : : : : : : : : :  
 9637 TAACGGTACCAAAATGACTACCAAGACCTACAGCGAACTGTCCGTTACTA 9686  
 481 snProTyrPheThrValAsnGlyValSerGlnSerLeuSerGlyTyrTyr 497  
 ||||| : : : : : ||||| : : : : : : : : : : : : : : : : :  
 9687 ACCATACCTTTACCGTTGCGGGGTGAGCGTCCGCTGGTGTATCTTCAT 9736  
 498 ArgLysThrLysTyrAspAsnLysAsnIleSerAsnTyrValLeuAspSe 514  
 : : : : : ||||| : : : : : : : : : : : : : : : : : : : :  
 9737 AACGACTTTGAACGGATGATCCGACCTGTCCGACTATACCAACAGAG 9786  
 514 rTyrGlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnArgI 531  
 ||||| : : : : : ||||| : : : : : : : : : : : : : : : : :  
 9787 TTACGGTACGAGTGTGCGCTTGGCTTCCCATCAACGAGTACACACAGT 9836  
 531 leSerPheGlyLeuAsnAlaAspAsnThrLysLeuHisGlyGlyArgPhe 547  
 : : : : : ||||| : : : : : : : : : : : : : : : : : : : :  
 9837 TCGTGTGTTGGCTATGTCCATTAACAAACTG : : : : : : : : : : 9871  
 548 MetGlyIleSerAsnValLys...GlnLeuMetAlaAspGlyGlyLysI 563  
 ||||| : : : : : : : : : : : : : : : : : : : : : : : :  
 9872 .....TCCAACTGACCGCCGAGATCGCGATGGATCTTATCTTGA 9912

563 eGlnValAspAsnAsnGlyIleProAspPheLysHisAspTyrThrThr 580  
 : : : : : : : : : : ||||| : : : : : : : : : : : : : :  
 9913 ATCGATGGGCGACCCATCATCCAGCGATTTGCCGCTGACGACTTCACCT 9962  
 580 yrAsnAlaIleLeuGlyTyrAsnTyrSerSerLeuAspArgProValPhe 596  
 : : : : : ||||| : : : : : : : : : : : : : : : : : : : :  
 9963 TTAAAC.....TACGGCTGACCTATAACAAGCTTGACCGTGGTTATTTC 10006  
 597 ProThrGlnGly.....MetSerHisSerValAspLeuThrValGlyPh 611  
 ||||| : : : : : : : : : : : : : : : : : : : : : : : :  
 10007 CCACGGACGGCTCGCGAGTCAACCTGACGGTAAAGTGACGATCCAGG 10056  
 611 eGlyAspLysThrHisGlnLysValValTyrGlnGlyAsnIleTyrArgP 628  
 : : : : : : : : : : : ||||| : : : : : : : : : : : : : :  
 10057 TTCGACACGAATAGTACAAAGTGTGCTGGATACCGGACCTATGTGC 10106  
 628 roPheIleLys.....LysSerValLeuArgGlyTyrAlaLysLeuGly 642  
 ||||| : : : : : ||| : : : : : ||| : : : : : : : : : :  
 10107 CTATTGATAACGACCAAGTGGTGGTCTCGGGGCGTACTCGTTGGGT 10156  
 643 TyrGlyAsnAsnLeu.....ProPheTyrGluAsnPheTy 654  
 ||||| : : : : : ||| : : : : : ||||| : : : : : : : : : :  
 10157 TACGGCGATGTTTAGCGCGCAAGGAATCCGCTTCTATGAGAACTCTA 10206  
 654 rAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSerSerLeuGlyP 671  
 : : : : : ||||| : : : : : ||||| : : : : : : : : : : : : : :  
 10207 TCGCGTGGTTCAGTACCGTCCGTTCCAGTCCGATCGAATATATCGGCC 10256  
 671 roArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnGlnThrThr... 686  
 || : : : : : : : : : : : : : : : : : : : : : : : : : :  
 10257 CGAAAGCG.....GTCTATAAAATGGGCCACACTAGTTGG 10294  
 687 .....Le 687  
 10295 GACGATGATGATGATTACGAGACTGTACTCAGGAATCAGCGTGTAAATC 10344  
 687 uGlyGluValValGlyGlyAsnAlaLeuAlaThrPheGlySerGluLeuI 704  
 : : : : : ||||| : : : : : ||||| : : : : : : : : : : : : : :  
 10345 AGATGATCGGTAGGTGGTAAATGCCATGGCTGTCGCCAGCGCTGGAATTA 10394  
 704 leuProLeuProPheLysGlyAsp...TrpIleAspGlnValArgPro 719  
 || ||| : : : : : : : : : : : : : : : : : : : : : : : :  
 10395 TTACCCCGACGCGGTTTCATCAGCGAAAGTATGCCAACTCGTTCGTACC 10444  
 720 ValIlePheIleGluGlyGlnValPheAspThrThr...GlyMetAs 735  
 : : : : : ||||| : : : : : ||||| : : : : : : : : : : : :  
 10445 TCCTTCTTCGGGATATGGTACCGTTCGGATACGAACTGGGATCCGCT 10494  
 735 pLysGlnThrIleAspLeuThrGlnPheLysAspProGlnAlaThrAlaG 752  
 : : : : : ||||| : : : : : ||||| : : : : : : : : : : : :  
 10495 ATCTGCGCGCTCAGATGTACCGGATTACAGCGATCCA..... 10531  
 752 luGlnAsnAlaLysAlaAlaAsnArgProLeuLeuThrGlnAspLysGln 768  
 : : : : : : : : : : : : : : : : : : : : : : : : : :  
 10532 .....GCCAAC 10537  
 769 LeuArgTyrSerAlaGlyValGlyAlaThrTyrTyrThrProIleGlyPr 785  
 : : : : : ||||| : : : : : ||||| : : : : : : : : : : : :  
 10538 ATCCGATGTCCCGGGATTCGCATTTACATGGATGTCCCATTTGGGCC 10587  
 785 oLeuSerIleSerTyrAlaLysProLeuAsnLysLysGlnAsnAspGlnT 802  
 ||||| : : : : : ||||| : : : : : : : : : : : : : : : : :  
 10588 GTTGGTCTTCTCTACGCCAGCGGTTTAAAAAGTACGATGGAGACAAAG 10637  
 802 hrAspThrValGlnPheGlnIleGlySerValPhe 813  
 : : : : : ||||| : : : : : ||||| : : : : : : : : : : : :  
 10638 CCAGCAGTTCAGTTTAACTGGTAAACCTGG 10672

seq\_name: gb\_ba:AF407013

seq\_documentation\_block:

LOCUS AF407013

DEFINITION Escherichia coli outer membrane protein (ecfk) gene, complete cds.

linear BCT 10-SEP-2001

2700 bp

DNA



```

1131 CAAAAAGCTTCGCGTCGCTATGGTTATGCCATCCGCGGTACAGTCGA 1180
331 aThrArgIleAsnAspGluSerArgThrValAspValGluIle 347
1181 TCCCGCAAAATTACGATGCCGCAAAACGGTTAAATTACGTGTGAACGTT 1230
348 AspProValHisProValTyrValArgArgIleAsnPheThrGlyAsnPh 364
1231 GATCGGGTTAACCGTTTCTACGTGCGTAGATCCGTTTGAAGGTAACGA 1280
364 eLysThrGlnAspGluValLeuArgGluMetArgGlnLeuGluGlyA 381
1281 TACCTCGAAAGATGCCGTCCTCGCTCGCGAAATGCGTCAGATGGAAGTG 1330
381 laLeuAlaSerAsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArg 397
1331 CATGCTGGGGAGCGATCGTGGTATCGAGTAAAGGCGTGAATCGT 1380
398 ThrGlyPhePheLysHisValThrValAspThrArgProValProAsnSe 414
1381 CTGGGCTCTTTGAACACTGTCGATACCGATACCCAAACGTGTTCCGGGTAG 1430
414 rProAspGlnValAspValAsnPhValValGluGluGlnProSerGlys 431
1431 CCGGACCAAGTTGATGTCGCTACAAGGTAAGGAGCGCAACACCGGTA 1480
431 eSerThrIleAlaIleGlyTyrSerGlnSerGlyGlyValThrPheGln 447
1481 GCTTCAACTTTGGTATTGTTACGTTACTGAAAGTGGCGTGAGCTTCCAG 1530
448 PheAspValSerGlnAsnAsnPhMetGlyThrGlyLysHisValAsnAl 464
1531 GCTGTGTGCACGAGTAACCTGGTTAGGTACAGGTTATGCTGTTGGTAT 1580
464 aSerPheSerArgSerGluThrArgGluValTyrSerLeuGlyMetThrA 481
1581 CAACGGGACCAAAACGATTACACAGCCTATGCTGAACCTGTCGGTAACCA 1630
481 snProTyrPheThrValAsnGlyValSerGlnSerLeuSerGlyTyrTyr 497
1631 ACCCTACTTACCGTAGATGCGGTAAAGCTCGGTGCTGCTCTCTAT 1680
498 ArgLysThrLysTyrAspAsnLysAsnIleSerAsnTyrValLeuAspSe 514
1681 AATGACTTCCAGGACAGATGACCCCGCACCTGTCGACTATACCAACAGAG 1730
514 rTyrGlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnArgI 531
1731 TTATGGTACAGACGTACCGTTGGGCTTCCGATTAAACGAATAAATACGCG 1780
531 leSerPheGlyLeuAsnAlaAspAsnThrLysLeuHisGlyArgPhe 547
1781 TCGGTGACGCTCGGTGTATGTACATAAATCCCTG..... 1815
548 MetGlyIleSerAsnValLysGlnLeuMetAla..... 558
1816 .....TCCACATCAGCGCTCAGGTGCGATGTCGCGTTATCTGTA 1856
559 AspGlyGlyLysIleGlnValAspAsnGlyIleProAspPheLysH 575
1857 CTCTATGGGTGAACATCCGAGCACCTCTGATCAGGATAACAGCTTCAAAA 1906
575 isAspTyrThrThrAsnAlaIleLeuGlyTrpAsnTyrSerSerLeu 591
1907 CGGACGACTTACGTTCAAC.....TATGGTGGACCTATATCAACAGCTT 1950
592 AspArgProValPheProThrGlnGly.....MetSerHisSerValas 606
1951 GACCGTGTACTTCCGCACAGATGGTTCACGTGTCAACCTGACCGGTAA 2000
606 pLeuThrValGlyPheGlyAspLysThrHisGlnLysValValTyrGlnG 623

```

```

2001 ACTGACCATTCTCGATCGGATAACGAATACTACAAAGTCACGTTAGACA 2050
623 lyAsnIleTyrArgProPhe.....IleLysLysSerValLeuArgGly 637
2051 CGCGGACTTATGTGCCGATCGATGACGATCACAATGGGTGTCTCGGG 2100
638 TyrAlaLeuLysLeuGlyTyrGlyAsnAsnLeu.....Proph 649
2101 CGTACCCGCTGGGTTATGTGTATGTTAGCGCGCAAGAGATGCCGTT 2150
649 eTyrGluAsnPheTyrAlaGlyTyrGlySerValArgGlyTyrAspG 666
2151 CTACGAGAACTTCTATCGCGGTGTTCCAGCCGTCGCTGGCTCCAGT 2200
666 InSerSerLeuGlyProArgSer.....GlnAla..... 675
2201 CCAATACCATTTGTCGAAAGCAGTTTACTTCCCGCATCAGGCCAGTAAT 2250
676 .....TyrLeuThrAlaArgArgGlyGlnGlnThrTh 686
2251 TATGATCCGCGACTATGATTACGAATGTGCGACTCAGGACGCGCGAAGA 2300
686 rLeu.....GlyGluValValGlyGlyAsnAlaLeuAlaThrPheG 700
2301 CCGTGTAAATCGGATGCTGTAGCGGTAAACGCCATGGCGTTGCCA 2350
700 lySerGluLeuIleLeuProLeuProPheLysGlyAsp...TrpIleAsp 715
2351 GCCTCGAGTTTCATCACCCGCGCGCTTATTAGCGATAAGTATGCTAAC 2400
716 GlnValArgProValIlePheIleGlyGlyGlnValPheAspThrTh 732
2401 TCGGTTCTGACTCTCTCTCGGATATGGTACCGTTTGGGATACAA 2450
732 rGlyMetAspLysGlnThrIleAspLeuThrGlnPheLysAspProGlnA 749
2451 CTGGGATTCAGCCCAATATCTCGATATCCGACTATAGTATGCCA.... 2496
749 laThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuThrGln 765
2496 ..... 2496
766 AspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrTrpTyrThrPr 782
2497 ...AGCAATATCCGTATGCTCGGGTATCGCATTAATGGATGCCCC 2543
782 oileGlyProLeuSerIleSerTyrAlaLysProLeuAsnLysLysGlnA 799
2544 ATTGGGCGGTGGTGTCTCTACGCCGCGGTTCAAAAAGTACGATG 2593
799 snAspGlnThrAspThrValGlnPheGlnIleGlySerValPhe 813
2594 GAGACAAGCAGACAGTTCACGATTAACTCGTTAAAAACCTGG 2637

```

seq\_name: gb\_ba:AE005193

seq\_documentation\_block:

LOCUS AE005193 10649 bp DNA linear BCT 21-MAR-2001  
 DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 17

ACCESSION AE005193

VERSION AE005193.1 GI:12512893

KEYWORDS

SOURCE

ORGANISM

Escherichia coli O157:H7 EDL933.  
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

REFERENCE

1 (bases 1 to 10649)

Perla, N.T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J.D.,  
 Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A.,  
 Posfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L.,  
 Grotbeck, E.J., Davis, N.W., Lim, A., Lim, A., Dimalanta, E., Potamousis, K.,  
 Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,

TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 AUTHORS  
 Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.  
 Direct Submission  
 Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA  
 FEATURES  
 source  
 Location/Qualifiers  
 1. 10649  
 /organism="Escherichia coli O157:H7 EDL933"  
 /strain="EDL933"  
 /serotype="O157:H7"  
 /db\_xref="taxon:155864"  
 /note="enterohemorrhagic"  
 215. 940  
 /gene="rpsB"  
 /note="rpsB"  
 215. 940  
 /gene="rpsB"  
 /function="structural component; Macromolecule synthesis, modification; Ribosomal proteins - synthesis, modification; Ribosome"  
 /note="Residues 1 to 241 of 241 are 100.00 pct identical to residues 1 to 241 of 241 from Escherichia coli K-12 strain MGI655; B0169"  
 /codon\_start=1  
 /transl\_table=11  
 /product="30S ribosomal subunit protein S2"  
 /protein\_id="AAG54471.1"  
 /db\_xref="GI:12512894"  
 /translation="MATVSRMDLKAGVHGHOTVYNNPKMPPFIFGARKVHIINLE KTVPMNEALAEINKIASRKGIKLVGTRKAAAEAVKDAALSCDQFFVNNHRLVGGMLT NWTYRQSKIRLKDLETQSDGTFDKLTKEALMRTELEKLENSLGGIKDMGLDPA LFDVADHIEHIAKEANNLGIPIVFAIVDTNSDDPDGVDVFIPIGNDDAIRAVTLILGAVA ATVREGRSODLASQAEESEVEAE"  
 1198. 2049  
 /gene="tsf"  
 /note="tsf"  
 1198. 2049  
 /gene="tsf"  
 /function="factor; Macromolecule synthesis, modification; Proteins - translation and modification"  
 /note="Residues 1 to 283 of 283 are 100.00 pct identical to residues 1 to 283 of 283 from Escherichia coli K-12 strain MGI655; B0170"  
 /codon\_start=1  
 /transl\_table=11  
 /product="protein chain elongation factor EF-Ts"  
 /protein\_id="AAG54472.1"  
 /db\_xref="GI:12512895"  
 /translation="MAETASLVKELRRTGAGMDCCKKALTEANGDIELAIENMRKS GAIAKAKAGNVAADGVIKIDGNVGIILEYVNCQDTEFAKADGAFADKVLDAAVA GKITDVEVLKAQFEERVALVAGENINIRNVAALGEGDVLGYSQHGARGVLVAAG ADELVKHIAHVAASPEFIKPEDVSAEYVEKYEQVOLDIAMQSGKPKETAEKMEVG RMRKFTGTEVSLQPPFVMEPSKTVGLLKEHNAEVTGFIREFVGEIEKVTEDFAAEV AAMSKQS"  
 2196. 2921  
 /gene="pyrH"  
 /note="pyrH"  
 2196. 2921  
 /gene="pyrH"  
 /function="enzyme; Central intermediary metabolism; Nucleotide interconversions"  
 /note="Residues 1 to 241 of 241 are 100.00 pct identical to residues 1 to 241 of 241 from Escherichia coli K-12

Strain MGI655: B0171"  
 /codon\_start=1  
 /transl\_table=11  
 /product="uridylylate kinase"  
 /protein\_id="AAG54473.1"  
 /db\_xref="GI:12512896"  
 /translation="MATNAKPVYKRIILLKSGALOGTEGFGDASILDMAOEIKEL VELGIQGVVIGGNLFRGAGLAKAGMNRVVGDMGLATVMNGLAMRDALHAYVNA RUMSAIPLNGVCDSYSWAEBASILLRNNRVVILSAGTGNPFPTTDSACLRGITEADV RLKATVDGVTADPAKDTATMYEQITYSEVLEKELKVMDLAAFTLARDHKLPIRVF NNNKPCALRRVYMGKEGELITE"  
 3071. 3628  
 /gene="fir"  
 /note="fir"  
 3071. 3628  
 /gene="fir"  
 /function="factor; Macromolecule synthesis, modification; Proteins - translation and modification"  
 /note="Residues 1 to 185 of 185 are 100.00 pct identical to residues 1 to 185 of 185 from Escherichia coli K-12 strain MGI655; B0172"  
 /codon\_start=1  
 /transl\_table=11  
 /product="ribosome releasing factor"  
 /protein\_id="AAG54474.1"  
 /db\_xref="GI:12512897"  
 /translation="MISDIRKDAEVRMCKVEAFKQISKIRGRASPSLLDGIVVEY YGPTPLQLASVTVEDSRPLKINVFDRSPAVEAKAIMASDLGLNPSAGSDIRVPL PPTESRRKDLTKIVRGEAQARVAVNRVRDANKVKALLKKEISEDDDRRSQDIDV OKLTDAAIKIEALADKEALMQF"  
 3720. 4916  
 /gene="yaeM"  
 /note="yaeM"  
 3720. 4916  
 /gene="yaeM"  
 /function="putative transport; Not classified"  
 /note="Residues 1 to 398 of 398 are 99.74 pct identical to residues 1 to 398 of 398 from Escherichia coli K-12 Strain MGI655; B0173"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative ATP-binding component of a transport system"  
 /protein\_id="AAG54475.1"  
 /db\_xref="GI:12512898"  
 /translation="MKOLITLGTSGICSTLDVVRHNPFRVVALVAGKNVTRMVE QCLEFSPRYAVMDDEASAKLTKMLOOQSRTEVLGQQAACDMALEDVDQVMAAIV GAAGLLPTLAIIRAGKTILLANKESLVTGCRLEMDAVKQSKAOLLPVDSHNAIFQSL PQPIQHNLYADLEONGVVSILLTGSGGPFREPLRDLATMTPDQACRHPNWSMGRKI SVDSATMMNKLEYIEARWLFNASASQMEVLIHPQSVIHSWRYQDSVLAQLGEPDM RTPTAHTMWPNRVNSGVKPLDFCKLSALTFAAPDPDYRYPCLKLAMEAFEGQAATTA LNAANEITVAFLAAQQIRFTDIAALNLVLEKMDMREPQCVDVVDVLSVDSAREYARKE VMRLAS"  
 5102. 5863  
 /gene="yaeS"  
 /note="yaeS"  
 5102. 5863  
 /gene="yaeS"  
 /function="orf; Unknown function"  
 /note="Residues 1 to 253 of 253 are 100.00 pct identical to residues 1 to 253 of 253 from Escherichia coli K-12 strain MGI655; B0174"  
 /codon\_start=1  
 /transl\_table=11  
 /product="orf, hypothetical protein"  
 /protein\_id="AAG54476.1"  
 /db\_xref="GI:12512899"  
 /translation="MMLSATQPLSEKLPAGCCRHVAIIMDNGEWAQKQKIRAFGHK AGKSVRRVAFSANNNGIEALTLYAFSENWNRPAQEVSALMELFVWALDSEVSLRHE HNVRLRIIGDTSRFNSRLQRIKSEALTAGTGLTGLTNIANYGGWIDVGVGQVQLAE KVOQGNLQPDQIDDEMLNQHVCMHELAPVDLVIIRTGGEHRISNLFLLQWIAEYLFYTD VLWPDFEQDQFEFEGALNFAFRERRRFGGTGPGDETA"  
 5984. 6733

gene

```

/gene="cdsA"
/feature="enzyme; Fatty acid biosynthesis: Fatty acid and
phosphatidic acid biosynth"
/feature="Residues 1 to 249 of 249 are 100.00 pct identical
to residues 1 to 249 of 249 from Escherichia coli K-12
Strain MG1655; B0175"
/codon_start=1
/trans_table=11
/product="CDP-diglyceride synthetase"
/protein_id="AAG54477.1"
/db_xref="GI:12512900"
/translation="MLAAEWGQLSGFTTSRVSRLAVLGLLLALMLFLPEVHRNI
HOPLVESLSASIGWIVALLLVLFYFGSAIWRNSKTLRLIFGLVLIIVFFWGMAL
RAWHDHYSGATLWLYMLVWGADSGAYMFKLFGKHKLPKVPSPGKTWQGFIG
LATAAIVSWGGMWANDVAPVTLLICISIVAALASVLGDTESMFKREAGIKDSGHLI
PGHGIIDRIDSLTAAVVFVACLLLLVFTL"
6745..8097
/gene="yael"
/feature="orf; Unknown function"
/feature="Residues 1 to 450 of 450 are 100.00 pct identical
to residues 1 to 450 of 450 from Escherichia coli K-12
Strain MG1655; B0176"
/codon_start=1
/trans_table=11
/product="orf, hypothetical protein"
/protein_id="AAG54478.1"
/db_xref="GI:12512901"
/translation="MISFLMDLASFIVAGLVLTNHEFGHFWARRCGVVERESIGF
GKALWRTDKLGEYVIALPLGQYVXMLDERAPVPELHNFNKNKSVQQRALIA
AGPVAFIFALFWLYIIGVPGVPVVGGEIAANSIAEAQIAPGTGLKAVDGIETP
DMDAVRLQDKIGDETTITVAFPGSDQRDLDRHMAFEDPKEDPVSSLGIRPR
GPQIEPVLENQVNSAASKAGLQAGDRIVKVDGQPLTQWTVFVLMVRDNPFGKSLALEI
alignment_scores:
Quality: 1212.00 Length: 848
Ratio: 2.200 Gaps: 15
Percent Similarity: 64.976 Percent Identity: 33.844
alignment_block:
US-09-701-711-2 x AE005193
Align seg 1/1 to: AE005193 from: 1 to: 10649
10 GlnValSerAlaMetThrMetAlaValMetMetValMetSer..... 23
8115 CGCATAATAACGATGCGATGAAAAAGTTGCTCATAGCGTCGCTGCTGT 8164
24 .....ThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIleT 37
8165 TAGCAGCGCCACGATACGGTGTGAGGGTTCGTAGTGAAGATATTC 8214
37 hrIleThrGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeu 53
8215 ATTTCGAAGCCCTTACGCTGTGCGGTGGTGGCGCCCTCTCAGTAG 8264
54 ProPheArgLeuGlyGlnValValSerGluAsnGlnLeuAlaAspGlyVa 70
8265 CCGGTGCGCACAGGCACACGGTTAATGATGAAGATATCATCAGTAATACCAT 8314
70 llysAlaLeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisG 87
8315 TCGCGCTGTGTTGTACCGCACTTTGAGGATGTTCCGCTCTCGTG 8364
87 lngluGlyArgIleIleTyrGlnValThrGluArgProLeuIleAlaGlu 103
8365 ATGGTGATACCTTCGTGGTTCAGGTAAAGAAGACGCCGACCATTCGCCACC 8414

```

```

104 IleAsnPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluG 120
111:::111:::111:::111:::111:::111:::111:::111:::111:::
8415 ATTACTTTCTCGGTAAACAAATCGGTGAAGATGACATGCTGAAGCAAAA 8464
120 yLeuLysAsnAlaGlyLeuAlaValGlyGlnProLysGlnAlaThrV 137
111:::111:::111:::111:::111:::111:::111:::111:::111:::
8465 CCTCAGGAGTTCGTGTGCGTGGCGAATCCCTCGATCGCACCAACA 8514
137 aGlnMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyr 153
111:::111:::111:::111:::111:::111:::111:::111:::111:::
8515 TTGCGGATATCGAGAAAGGTTCGGAAGACTTCTACAGCGTCGCTGATAA 8564
154 TyrAsnThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgVa 170
111:::111:::111:::111:::111:::111:::111:::111:::111:::
8565 TATAGCGCCAGCGTTAAAGCTGCTGACCCGCTGCCGCGCAACCGTGT 8614
170 llysLeuAspMetThrPheAlaGlyLysProAlaArgValValAspI 187
111:::111:::111:::111:::111:::111:::111:::111:::111:::
8615 TGACCTAAACTGGTGTTCAGGAAGGTGTGTCAGCTGAAATCCAGCAAA 8664
187 leaSnIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspVal 203
111:::111:::111:::111:::111:::111:::111:::111:::111:::
8665 TTAACATTGTTGTAACCATGCTTCCACCAGCAGCAACTGATCTCTCAT 8714
204 LeuAlaIleLysAsp.....AsnLysIleAsnProLeuSe 215
111:::111:::111:::111:::111:::111:::111:::111:::111:::
8715 TTCCAACCTGCGTACCAAGTGGTGGTGAAGCTGTA..... 8753
215 rLysAlaAspArg...TyrThrGlnGlyLysLeuValThrSerLeuGluA 231
111:::111:::111:::111:::111:::111:::111:::111:::111:::
8754 ....GGCGATCGTAAATACCAAGAAACAGAACTGGCGGCGACCTTGAAA 8799
231 snLeuArgAlaLysTyrLeuAsnAlaGlyPheValArgPheGluIleLys 247
111:::111:::111:::111:::111:::111:::111:::111:::111:::
8800 CCTGCGCAGCTACTATCTGGACCGCGGTATGCCCGCTTTCAACATAGAC 8849
248 AspAlaLysLeuAsnIleAsnGluAspLysAsnArgIlePheValGluI 264
111:::111:::111:::111:::111:::111:::111:::111:::111:::
8850 TCTACCCAGGTGAGTCTGACGCGCAGATAAAAAAGGTATTTCACGTACGGT 8899
264 eSerLeuHisGlyGluGlnTyrArgPheGlyGlnThrGlnPheLeuG 281
111:::111:::111:::111:::111:::111:::111:::111:::111:::
8900 GAACATCACCGAAGCGGATGACGTACAAAGCTTCTGCGGTGGAAGTGAGCG 8949
281 lYasnLeuThrTyrThrGlnAlaGluLeuAlaLeuLeuLysPheLys 297
111:::111:::111:::111:::111:::111:::111:::111:::111:::
8950 GCAACCTTCGCGGCACCTCGCTGAAATTCAGCAGCTGACTAAGATCGAG 8999
298 AlaGluGluGlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsnI 314
111:::111:::111:::111:::111:::111:::111:::111:::111:::
9000 CCGGCGAGCTGTATAACGCGCACCAAGTGCACCAAAATGGAAGATGACAT 9049
314 eSerThrLysPheGlyAspGlyTyrTyrTyrAlaGlnIleArgProV 331
111:::111:::111:::111:::111:::111:::111:::111:::111:::
9050 TAAAAAGCTTCGCGTGCATGGTATGCTATCCGCGCGGTACACGTCGA 9099
331 alThrArgIleAsnAspGluSerArgThrValAspValGluTyrTyrIle 347
111:::111:::111:::111:::111:::111:::111:::111:::111:::
9100 TCGCTGAAATTAACGATGCGCAACAAACCGTTAAATTCGCGTGAACGTT 9149
348 AspProValHisProValTyrValArgIleAsnPheThrGlyAsnPh 364
111:::111:::111:::111:::111:::111:::111:::111:::111:::
9150 GATGCGGTAAACGCTTTCACGTGCGTAAAGTCCGTTTGAAGGTAAACGA 9199
364 elysThrGlnAspGluValLeuArgGluMetArgGlnLeuGluGlyA 381
111:::111:::111:::111:::111:::111:::111:::111:::111:::
9200 TACCTCGAAAGATGCGCTCGTGGCAAAATGCGTGCAGATGGAAGGTG 9249
381 laLeuAlaSerAsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArg 397
111:::111:::111:::111:::111:::111:::111:::111:::111:::
9250 CATGGCTGGAGCGATCTGCTGATCAGGTAAGGAGCGTCTGAATCGT 9299
398 ThrGlyPhePheLysHisValThrValAspThrArgProValProAsnSe 414

```





## JOURNAL

Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using GeneMark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA.

30332 [e-mail: markamber.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (<http://cgsc.biology.yale.edu>). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (<http://www.genetics.wisc.edu>). \*\*\* The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

## FEATURES

## Location/Qualifiers

```

1..14877
  /organism="Escherichia coli K12"
  /strain="K12"
  /sub_strain="MG1655"
  /db_xref="taxon:83333"
123..884
  /gene="yaeS"
  /note="b0174"
123..884
  /gene="yaeS"
  /function="orf; Unknown"
  /note="O253; 61 pct identical to 229 residues of an approx.240 aa hypothetical protein Y920_HAEIN SW: P44938"
  /codon_start=1
  /transl_table=11
  /product="orf, hypothetical protein"
  /protein_id="AAC73285.1"
  /db_xref="GI:1786371"
  /translation="MMLSATQPLSEKLPAGCRHVAIIIMDNGRWAKQKIRAFGHK
  AGAKSVRAVFAANNQTEALTLYAFSSNNRPAQVSALEMLFVWALDSEVKSLSHR
  HNVRLITGDTSRFSRQERIRKSEALTAGNTGLTLNIAANYGGRWDIVQGVQLAE
  KVOQNLDPDIDIELMLNQHVMHELAPVDLIVRTGGEHRISNELLWQIAELYFTD
  VLWPDFRQDFEGALNAPNRERRGGTEPGEFA"
870..900
  /note="factor Sigma70; predicted +1 start at 195687"
899..927
  /note="factor Sigma70; predicted +1 start at 195714"
956..987
  /note="factor Sigma70; predicted +1 start at 195774"
1005..1754
  /gene="cdsA"
  /EC_number="2.7.7.41"
  /function="enzyme; Fatty acid and phosphatidic acid biosynthesis"
  /note="O249; 100 pct identical to CDSA_ECOLI SW: P06466"
  /codon_start=1
  /transl_table=11
  /product="CDP-diglyceride synthetase"
  /protein_id="AAC73286.1"
  /db_xref="GI:1786372"

```

## protein\_bind

```

/translation="MLAAEWGQLSGFTTRSORVLAIVLGLLLALLMLFLPLPEYHRNI
HQPLVEISLWASLGHVIVALLLVLPYPGSAAIWRNSKILRIFGVITIVPFPFGMLAL
RAWHYDENHSGAIIWLLVYVILLVWGADSGAYMFGKLPGRKHLKAPVSGKTRWQGTIGG
LATAVAISWGYGMWNLVDVAPVTLLICISIVAAALSVGLDITESTMFREKAGIKDSGHLI
PGHGILDRIDSLTAAPVPVFACLLLLVFRTL"
1421..1432
  /gene="cdsA"
  /note="No predicted promoter"
  /bound_moiety="PuA predicted site"
1766..3118
  /gene="yaeL"
  /note="b0176"
1766..3118
  /gene="yaeL"
  /function="orf; Unknown"
  /note="O450; 100 pct identical to fragment YAEI_ECOLI
  SW:P37784 but contains 302 additional C-terminal residues"
  /codon_start=1
  /transl_table=11
  /product="orf, hypothetical protein"
  /protein_id="AAC73287.1"
  /db_xref="GI:1786373"
  /translation="MLSFLMLASFIVALGVLTIVHERGHFWARCGVRVERFSTIGF
  GKALWRTDKLGTETEVIALPLGGVYKMLDERAEVPELHAFHNNKSVGQRAAIIA
  AGPVANFIFAIVLWVLTIGVGVVPPVGEIAANSAIAEAQIAPGTCLKAVDGIETP
  DWDVRLQLVDKIGDESTITVAPFGSDQRDVKDLDRHWAPEPKEDPVSSIGIRPR
  GPQIVLENVQPNASAASKAGIQAQDRIVKVDGQPLTQWTFVMLVRPNQPSLSALEI
  ERQSGSLTILPESKPGNGKAIGFVIEPKVILPDEKYKVRQGPFNALVEADTKT
  WOLMKLTVMIGLKLTIGDVKNLNSGPISIAKGAGMTAELGVVYLYPLFALLISVNLGI
  INLFPLVLDGGHLLFLATEIKTGGPSERVQDFCYRIGSILLVLLMLGLALNDFESRL"
3148..5580
  /gene="yaeR"
  /note="b0177"
3148..5580
  /gene="yaeR"
  /function="orf; Unknown"
  /note="O810; 45 pct identical (29 gaps) to 808 residues of
  an approx. 800 aa protein D151_HAEIN SW: P46024"
  /codon_start=1
  /transl_table=11
  /product="orf, hypothetical protein"
  /protein_id="AAC73288.1"
  /db_xref="GI:1786374"
  /translation="MAMKLLIASLFLSSATVYGAEGFVKDIHFEGLQVAVGAALL
  SMPVPTGPTVNDSDISNTIRALFATGNEEDYVLPDGTLLVQVKERTPIASITFSGN
  KSVKDDMLKQNLASGVVGVESLDRITTIADIEKLEDFYISVGKYSASVAVTFLPR
  NRVDLKLVFQEGVSAEQIINIVGNHAFITDELISHFQLRDVEPVMNVVGGRRYQKQK
  LAGDLETLSYLDRGYARFNIDSTQVSLTPDKGIIYTVNITEGDTYKLSGVEYSGN
  LAGSAETEQLTKEPGEYNGTKVTKMEDDIKLLGRYAYPRVQSMPEINDADKT
  VKLRVNDAGNRVYVRKTRFEGNDTSKDAVLRREMRQMEGAWLGSGLVDQGERLNRL
  GFFETVDTQVPCSPQDVVYKVKERTGSGNFSGIGYCTESGVQAGVQOQDNL
  GTGAVGLNGRNDQYQTELVSINPYFTVDGSLGRFLYNDFAADLSDYTNKS
  YGTDVTGLFPINENSLRAGLSVHNSLNMQPQVAMRRLYSMEHFSTSDQNSFK
  TDDTFNYGWTYNKLDRGYFPTDGRVNLTKGVTIPGSDNYKYVTLTATVYPTDDO
  HKWVVLGRTRWYGDIGGKPEPFYFVAGSSTVGVQSNITGPKAVYFPHQASNY
  DPDYDECATODGAKDLCKSDDAVGSNAWASLEFITPTPEISPKYANSVTSFTFWD
  MGTVDVNDWSQYSGYDYPDSNIRMSAGIALQWMSPLGLPVFSYAQPFFKYDGDK
  AEQFQFNIGKTW"
5583..5611
  /note="factor Sigma70; predicted +1 start at 200398"
5702..6187
  /gene="hlpA"
  /note="b0178"
5702..6187
  /gene="hlpA"
  /function="factor; Basic proteins - synthesis,
  modification"
  /note="O161; 100 pct identical to HLP_A_ECOLI SW:
  P11457; similar to ompH of other bacterial species"
  /codon_start=1
  /transl_table=11
  /product="histone-like protein, located in outer membrane

```

## promoter

## gene

## CDS

```

or nucleoid"
/protein_id="AAC73289.1"
/db_xref="GI:1786375"
/translation="MKKVLAAAGLGLALATSAQASDRIATVNMGSLQFQVQAQGTGVS
TLNEFKGASLQETMDLQAKMKLQSMAGSDRDKLERYDAQQRTPFAQKAQAFE
QDARRSNRNEGRGLVTRIQTAVKSVANSQDIDLVDANAVAYNSSDVKIDITADVLRQV
K"

```

gene

CDS

alignment\_scores:

Quality:	1212.00	Length:	848
Ratio:	2.200	Gaps:	15
Percent Similarity:	64.976	Percent Identity:	33.844

alignment block:

US-09-701-711-2 x AE000127

Align seq 1/1 to: AE000127 from: 1 to: 14877

10 GlnValSerAlaMetThrMetAlaValMetMetValMetSer..... 23  
 3136 CGCATAAACGATGGCGATGAAAAGTGTGCTACATGCGTGCCTGCTTT  
 24 .....ThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIle 37  
 3186 TAGCAGCGCCACCGCTATACGGTGTGAAGGTTGCTAGTGAAGATATTC 3235  
 37 hrIleThrGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeu 53  
 3236 ATTTCGAAGCGCTTCAGCGTGTCCCGTGTGGCGGCCCTCCTCAGTATG 3285  
 54 ProPheArgLeuGlyGlnValValSerGluAsnGlnLeuAlaAspGlyVa 70  
 3286 CCGGTGCCACAGCGCACCGGTTAAATGATGAAGATATCAGTAAATACAT 3335  
 70 LlysAlaLeuTyrrAlaThrGlyAsnPheSerAspValGlnValHisG 87  
 3336 TCGCGCTCTGTTGCTACCGGCACTTTGAGGATGTTGCGGCTCTCGTG 3385  
 87 lngLuGlyArgIleIleTyrrGlnValThrGluArgProLeuIleAlaGlu 103  
 3386 ATGTGTGATACCTTCTGGTTCAGGTAAAGAACGTCGACCATTCGCACG 3435  
 104 IleAsnPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGlu 120  
 3436 ATTACTTCTCCGGTAAACAAATCGGTGAAGATGACATGCTGAAGCAAAA 3485  
 120 yLeuLysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrV 137  
 3486 CCTCGAGCGTCTTGGTGTGGGTGGCGAATCCTCGATCGCACCAACCA 3535  
 137 aIlgmMetIleGluThrGluLeuThrAsnGlnTyrrIleSerGlnGlyTy 153  
 3536 TTGCGCATATCGAAGAGGTCTCGAAGACTTCTACTACAGCGTCGGTAAA 3585  
 154 TyrAsnThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgVa 170  
 3586 TATACGCCGCGAATAAAGTGTGTCGACCCCGCTGCCGCGACCGGTGT 3635  
 170 LlysLeuAspMetThrPheAlaGluGlyLysProAlaArgValValAspI 187  
 3636 TCACATAAACTGGTGTTCAGGAAGAGTGTGTCAGCTGAATCCAGCAAA 3685

```
|||||..... 5436
5471 ACCCGTACTTACCGTAGATGGCGTAAAGCTCGTCTCTCTCTAT 4620
498 ArgLysThrLysTyrAspAsnLysAsnIleSerAsnTyrValLeuAspSe 514
5437 ...AGCAATATCGTATGCTCGGGTATCGCAATACATGGATGCC 5483
782 olleGlyProLeuSerIleSerTyrAlaLysProLeuAsnLysLysGlnA 799
5484 ATTGGGCGGTGGTGTCTCTCTACGCCAGCGCTTCAAAAAGTACGATG 5533
799 snAspGlnThrAspThrValGlnPheGlnIleGlySerValPhe 813
5534 GAGACAAGGCAGCAACAGATTCCAGTTTAACATCGGTAAACCTGG 5577
seq_name: gb_ba:AE008705
seq_documentation_block:
LOCUS AE008705 22997 bp DNA linear BCT 25-OCT-2001
DEFINITION Salmonella typhimurium LT2, section 13 of the complete
genome.
ACCESSION AE008705 AE006468
VERSION AE008705.1 GI:16418723
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Coding sequences below are predicted from manually evaluated
computer analysis, using similarity information and the programs;
GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and
GeneMark; http://opal.biology.gatech.edu/GeneMark/
EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/
and Pedro Romero and Peter Karp at Ecocyc;
http://ecocyc.PangeaSystems.com/ecocyc/
The analyses of ribosome binding sites and promoter binding sites
were kindly provided by Heladia Salgado, Julio Collado-Vides and
RequonDB;
http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset
This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate
chemistries or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one m13 subclone.
FEATURES
source
1. 22997
/organism="Salmonella typhimurium LT2"
/strain="LT2; SGSC 1412; ATCC 700720"
/db_xref="ATCC:700720"
/db_xref="taxon:99287"
98. .103
RBS
```

```
|||||..... 5436
5471 ACCCGTACTTACCGTAGATGGCGTAAAGCTCGTCTCTCTCTAT 4620
498 ArgLysThrLysTyrAspAsnLysAsnIleSerAsnTyrValLeuAspSe 514
5437 ...AGCAATATCGTATGCTCGGGTATCGCAATACATGGATGCC 5483
782 olleGlyProLeuSerIleSerTyrAlaLysProLeuAsnLysLysGlnA 799
5484 ATTGGGCGGTGGTGTCTCTCTACGCCAGCGCTTCAAAAAGTACGATG 5533
799 snAspGlnThrAspThrValGlnPheGlnIleGlySerValPhe 813
5534 GAGACAAGGCAGCAACAGATTCCAGTTTAACATCGGTAAACCTGG 5577
seq_name: gb_ba:AE008705
seq_documentation_block:
LOCUS AE008705 22997 bp DNA linear BCT 25-OCT-2001
DEFINITION Salmonella typhimurium LT2, section 13 of the complete
genome.
ACCESSION AE008705 AE006468
VERSION AE008705.1 GI:16418723
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Coding sequences below are predicted from manually evaluated
computer analysis, using similarity information and the programs;
GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and
GeneMark; http://opal.biology.gatech.edu/GeneMark/
EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/
and Pedro Romero and Peter Karp at Ecocyc;
http://ecocyc.PangeaSystems.com/ecocyc/
The analyses of ribosome binding sites and promoter binding sites
were kindly provided by Heladia Salgado, Julio Collado-Vides and
RequonDB;
http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset
This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate
chemistries or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one m13 subclone.
FEATURES
source
1. 22997
/organism="Salmonella typhimurium LT2"
/strain="LT2; SGSC 1412; ATCC 700720"
/db_xref="ATCC:700720"
/db_xref="taxon:99287"
98. .103
RBS
```



264 eSerLeuHisGluGlyGluInThrArgPheGlyGlnThrGlnPheLeuG 281  
: : : : : | | | | | : : : : :  
6115 CAATATCACCGAAGCATCACTACAAAGTTTCGGGGTTCAGGTGAGCG 6164

281 lValAsnLeuThrTyThrGlnAlaGluLeuGluAlaLeuLeuPheLys 297  
| | | | | : : : : : | | | | | : : : : :  
6165 GTAACCTGCCGGCATTCCGCTGAATTTGAACTTGACACTGCAGAAATTTGAA 6214

298 AlaGluGluGlySerGlnAlaMetLeuGluGlnThrThrAsnAsnI 314  
| | | : : : : : : : : : : : : : : :  
6215 CCGGGCAGCTATAACGGCACCAAAGTACCAAAATTTGGAAGATGACAT 6264

314 eSerThrLysPheGlyAspAspGlyTyThrTyraGlnIleArgProv 331  
: : : : : | | | | | : : : : :  
6265 CAAAAAGCTTCTGGGTGCTATGTTACGCTCATCCGCGCTGCAGTCAC 6314

331 alThrArgIleAsnAspGluSerArgThrValAspValGluTyTyri 347  
: : : : : | | | | | : : : : : | | | | | : : : : :  
6315 AGCCTGAATTAACGATGCTGACAAAACCGTAAACTGCGCGTAAACGTC 6364

348 AspProValHisProValTyValArgArgIleAsnPheThrGlyAsnPh 364  
| | | : : : : : | | | | | : : : : :  
6365 GATCGGGCAACCGTTCTACGTGGTAAGATCCGCTTTGAAGCAACA 6414

364 eLysThrGlnAspGluValLeuArgArgGluMetArgGlnLeuGluLy 381  
: : : : : | | | | | : : : : :  
6415 CACCTCAAAAGATTCCGTTCTGCGCGTGAATTCGCCAGATGGAAGGG 6464

381 laLeuAlaSerAsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArg 397  
| | : : : : : : : : : : : : : : :  
6465 CGTGCTGGGCAGCACCTGTTGACAGGGGAAGAGCGTCTCAACCGT 6514

398 ThrGlyPhePheLysHisValThrValAspThrArgProValProAsnSe 414  
: : : : : | | | : : : : : | | | | | : : : : :  
6515 CTCGGCTCTTTGAACCGTCGACACCGACGACGCGCGTTCGCGGTAG 6564

414 rProspGlnValAspValAsnPheValValcLuGlnProSerGlys 431  
: : : : : | | | | | : : : : : | | | | | : : : : :  
6565 CCGGATCAGGTTGATGGTGTACAGGTGAAAGAGCGCAACACCGTA 6614

431 erSerThrIlealaGlytyrSerGlnSerGlyGlyValThrPheGln 447  
| | : : : : : | | | | | : : : : : | | | | | : : : : :  
6615 GCTTTAAC TTCGGTTCGGTTCAGGTACAGAAAGCGCGTTAGTTCAC 6664

448 PheAspValSerGlnAsnPheMetGlyThrGlyLysHisValasna 464  
: : : : : | | | : : : : : | | | | | : : : : :  
6665 GCGCGGTTACAGCAGCACTGGTTAGTACCGGTATTCCTCGGGGAT 6714

464 aSerPheSerArgSerGluThrArgGluValTySerLeuGlyMetThrA 481  
: : : : : : : : : : : : : : : | | | : : : : :  
6715 TAACGGCACAAAACGACTACCACTACAGCAACTGTCCTTACTA 6764

481 snProtyrPheThrValasnGlyValSerGlnSerLeuSerGlyTyTr 497  
| | | | | : : : : : | | | | | : : : : : | | | : : : : :  
6765 ACCCATACTTTACGCTTACGCGGTGAGCCCTCGGTGCTGCTATCTAT 6814

498 ArgLysThrLysTyraAspAsnLysAsnIleSerAsnTyValLeuAspSe 514  
: : : : : | | | : : : : : | | | | | : : : : :  
6815 AAGACTCCCGGGCAGCAGCCGGATCTATCCGACTATACCAACAAG 6864

514 tTyrglyGlySerLeuSerTygTyPrOileaspGluAsnGlnArgI 531  
| | | | | : : : : : | | | | | : : : : : | | | : : : : :  
6865 TTAGGTACGGATGTGACGCTTGGCTTCCCCTCAACAGAGTACAACAGT 6914

531 leSerPheGlyLeuAsnAlaAspAsnThrLysLeuHisGlyGlyArgPhe 547  
: : : : : | | | | | : : : : : | | | | | : : : : :  
6915 TCGTGCTGCTGGGTATGTCCATAACAACATG..... 6949

548 MetGlyIleSerAsnValLys... GlnLeuMetAlaasp..... 559  
| | | | | : : : : : | | | | | : : : : :  
6950 .....TCCAACTACGCGCGCAGATCGCGATCGTTCATCTGTA 6990

560 .....GlvGlyLysIleGlnValAspAsnGlnGlyIleProaspPheL 574



```

TITLE      Nucleotide sequence of the xanthine guanine phosphoribosyl
JOURNAL    transferase gene of E. coli
MEDLINE    Nucleic acids research. 11 (24), 8809-8816 (1983)
PUBMED     84169517
REFERENCE  6324102
AUTHORS    Pratt,D. and Subramani,S.
TITLE      Nucleotide sequence of the Escherichia coli xanthine-guanine
JOURNAL    phosphoribosyl transferase gene
MEDLINE    Nucleic acids research. 11 (24), 8817-8823 (1983)
PUBMED     84169518
REFERENCE  6324103
AUTHORS    Deutch,A.H., Rushlow,K.E. and Smith,C.J.
TITLE      Analysis of the Escherichia coli proBA locus by DNA and protein
JOURNAL    sequencing
MEDLINE    Nucleic acids research. 12 (15), 6337-6355 (1984)
PUBMED     84297232
REFERENCE  6089111
AUTHORS    Jagadeeswaran,P., Ashman,C.R., Roberts,S. and Langenberg,J.
TITLE      Nucleotide sequence and analysis of deletion mutants of the
JOURNAL    Escherichia coli gpt gene in plasmid pSV2 gpt
MEDLINE    Gene. 31 (1-3), 309-313 (1984)
PUBMED     85128454
REFERENCE  6396164
AUTHORS    12 (bases 1 to 91430)
TITLE      Nuesch,J. and Schumperli,D.
JOURNAL    Structural and functional organization of the gpt gene region of
MEDLINE    Escherichia coli
PUBMED     Gene. 32 (1-2), 243-249 (1984)
REFERENCE  85155481
AUTHORS    13 (bases 1 to 91430)
TITLE      Becerrii,B., Valle,F., Merino,E., Riba,L. and Bolivar,F.
JOURNAL    Repetitive extragenic palindromic (REP) sequences in the
MEDLINE    Escherichia coli gdhA gene
PUBMED     Gene. 37 (1-3), 53-62 (1985)
REFERENCE  86031366
AUTHORS    14 (bases 1 to 91430)
TITLE      Nomura,T., Alba,H. and Ishihama,A.
JOURNAL    Transcriptional organization of the convergent overlapping dnaQ-rnh.
MEDLINE    genes of Escherichia coli
PUBMED     The Journal of biological chemistry. 260 (11), 7122-7125 (1985)
REFERENCE  85207736
AUTHORS    15 (bases 1 to 91430)
TITLE      Icho,T., Sparrow,C.P. and Raetz,C.R.
JOURNAL    Molecular cloning and sequencing of the gene for CDP-diglyceride
MEDLINE    synthetase of Escherichia coli
PUBMED     The Journal of biological chemistry. 260 (22), 12078-12083 (1985)
REFERENCE  86008268
AUTHORS    16 (bases 1 to 91430)
TITLE      Cox,E.C. and Horner,D.L.
JOURNAL    DNA sequence and coding properties of mutD(dnaQ) a dominant
MEDLINE    Escherichia coli mutator gene
PUBMED     Journal of molecular biology. 190 (1), 113-117 (1986)
REFERENCE  87060973
AUTHORS    17 (bases 1 to 91430)
TITLE      Takano,K., Nakabeppu,Y., Maki,H., Horiuchi,T. and Sekiguchi,M.
JOURNAL    Structure and function of dnaQ and mutD mutators of Escherichia
MEDLINE    coli
PUBMED     Molecular & general genetics : MGG. 205 (1), 9-13 (1986)
REFERENCE  87089093
AUTHORS    18 (bases 1 to 91430)
TITLE      Richardson,K.K., Richardson,F.C., Crosby,R.M., Swenberg,J.A. and
JOURNAL    Skopek,T.R.
MEDLINE    DNA base changes and alkylation following in vivo exposure of
PUBMED     Escherichia coli to N-methyl-N-nitrosourea or N-ethyl-N-nitrosourea
REFERENCE  3540531
AUTHORS
TITLE

```

---

```

JOURNAL    Proceedings of the National Academy of Sciences of the United
MEDLINE    States of America. 84 (2), 344-348 (1987)
PUBMED     87092406
REFERENCE  3540961
AUTHORS    19 (bases 1 to 91430)
TITLE      Horiuchi,T., Nagasawa,T., Takano,K. and Sekiguchi,M.
JOURNAL    A newly discovered tRNA(Iasp) gene (aspv) of Escherichia coli K12
MEDLINE    Molecular & general genetics : MGG. 206 (2), 356-357 (1987)
PUBMED     87228334
REFERENCE  3295485
AUTHORS    20 (bases 1 to 91430)
TITLE      Tomasiewicz,H.G. and McHenry,C.S.
JOURNAL    Sequence analysis of the Escherichia coli dnaE gene
MEDLINE    Journal of bacteriology. 169 (12), 5735-5744 (1987)
PUBMED     88058791
REFERENCE  3316192
AUTHORS    21 (bases 1 to 91430)
TITLE      Crowell,D.N., Reznikoff,W.S. and Raetz,C.R.
JOURNAL    Nucleotide sequence of the Escherichia coli gene for lipid A
MEDLINE    disaccharide synthase
PUBMED     Journal of bacteriology. 169 (12), 5727-5734 (1987)
REFERENCE  88058790
AUTHORS    22 (bases 1 to 91430)
TITLE      Coleman,J. and Raetz,C.R.
JOURNAL    First committed step of lipid A biosynthesis in Escherichia coli:
MEDLINE    sequence of the lpxA gene
PUBMED     Journal of bacteriology. 170 (3), 1268-1274 (1988)
REFERENCE  88139188
AUTHORS    23 (bases 1 to 91430)
TITLE      Henrich,B., Schroeder,U., Frank,R.W. and Plapp,R.
JOURNAL    Accurate mapping of the Escherichia coli pepD gene by sequence
MEDLINE    analysis of its 5' flanking region
PUBMED     Molecular & general genetics : MGG. 215 (3), 369-373 (1989)
REFERENCE  89218945
AUTHORS    24 (bases 1 to 91430)
TITLE      Ichikawa,S. and Kaji,A.
JOURNAL    Molecular cloning and expression of ribosome releasing factor
MEDLINE    The Journal of biological chemistry. 264 (33), 20054-20059 (1989)
PUBMED     90062117
REFERENCE  2684966
AUTHORS    25 (bases 1 to 91430)
TITLE      Zhou,Z. and Syvanen,M.
JOURNAL    Identification and sequence of the drpA gene from Escherichia coli
MEDLINE    Journal of bacteriology. 172 (1), 281-286 (1990)
PUBMED     90094229
REFERENCE  1688424
AUTHORS    26 (bases 1 to 91430)
TITLE      Henrich,B., Monnerjahn,U. and Plapp,R.
JOURNAL    Peptidase D gene (pepD) of Escherichia coli K-12: nucleotide
MEDLINE    sequence, transcript mapping, and comparison with other peptidase
PUBMED     genes
REFERENCE  Journal of bacteriology. 172 (8), 4641-4651 (1990)
AUTHORS
TITLE

```

---

```

alignment_scores:
Quality: 1212.00      Length: 837
Ratio: 2.216         Gaps: 14
Percent Similarity: 65.352 Percent Identity: 34.050

alignment_block:
US-09-701-711-2 x ECOTSF ..
Align seg 1/1 to: ECOTSF from: 1 to: 91430
15 ThrMetalAlaValMetMetValMetSerThrHisAlaGlnAlaAlaAspPh 31
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7507 ACGGTGTCGCTGCTTTAGCAGCGCCGCCACCGTATACGGTGTCTGAAGGTT 7556
31 eMetAlaAsnAspIleThrIleThrGlyLeuGlnArgValThrIleGlu 48
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7557 CGTAGTGAAGATATTCATTTCGAAGCGCTTCAGCGTGTCCCGCTGGTG 7606

```













```
|||||
31096 CCTGTCTAAATCGGATGCTGTAGCGGTACGCCATCGCGGTGCCA 31145
      :::::::::::|||||
700 lyserylLeuLeuLeuProLeuProPheLysGlyAsp...TrpIleAsp 715
      :: |||:::|||| ||| ||||| ::||| ::|
31146 GCTCGAGTTCATCACCGCGCGCGCTTTATTAGCGATAAGTATGCTAAC 31195
      ::|||:::|||||
716 GlnValArgProValIlePheIleGluGlyGlyGlnValPheAspThr 732
      ::|||:::|||||
31196 TCGGTTCTGACTCTCTCTCTGGGATATGGTACCGTTTGGGATACAA 31245
      ::|||:::|||||
732 rGlyMetAspLysGlnThrIleAspLeuThrGlnPheLysAspProGlnA 749
      ::|||:::|||||
31246 CTGGGATTCAGGCATATCTCGGATATCCGGACTATAGTATGCCA... 31291
      ::|||:::|||||
749 laThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuThrGln 765
31291 ..... 31291
766 AspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrTriptyrThrPr 782
      ::|||:::|||||
31292 ...AGCAATATCCGTATCTGCTCGCGGTATCGCATTCATCATGATGCC 31338
      ::|||:::|||||
782 orLeGlyProLeuSerIleSerTyrAlaLysProLeuAsnLysLysGlnA 799
      ::|||:::|||||
31339 ATTGGGGCGCTGTCTCTCTACGCCCGCGCTTCAAAAAGTACGATG 31388
      ::|||:::|||||
799 snAspGlnThrAspThrValGlnPheGlnIleGlySerValPhe 813
      ::|||:::|||||
31389 GAGACAAAGCAGAACAGTTCACGTTTAACATCGGTAAACCTGG 31432
      ::|||:::|||||
seq_name: gb_ba:AP002550

seq_documentation_block:
LOCUS AP002550 281530 bp DNA linear BCT 07-MAR-2001
DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 1/20.
ACCESSION AP002550 BA000007
VERSION AP002550.1 GI:13359456
KEYWORDS
SOURCE Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
DNA
ORGANISM Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE
AUTHORS Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C.H., Kimura, S.,
Kurokawa, K., Ishii, K., Hattori, M., Tatsuno, I., Abe, H., Iida, T.,
Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T.,
Sasakawa, C., and Shinagawa, H.
complete nucleotide sequence of the prophage VT2-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)
20198780
JOURNAL
MEDLINE
AUTHORS Ohnishi, M., Murata, T., Nakayama, K., Kuhara, S., Hattori, M.,
Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shinagawa, H. and
Hayashi, T.
Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain M61655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
20557356
REFERENCE
AUTHORS Yokoyama, K., Makino, K., Kubota, Y., Watanabe, M., Kimura, S.,
Yutsudo, C.H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T.,
Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C., and
Shinagawa, H.
Complete nucleotide sequence of the prophage VT1-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the Sakai outbreak
Gene 258 (1-2), 127-139 (2000)
20564182
JOURNAL
MEDLINE
AUTHORS
```

4 (sites)  
Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K.,  
Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T.,  
Tanaka, M., Tobe, F., Iida, T., Takami, H., Honda, T., Sasakawa, C.,  
Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. and  
Shinagawa, H.  
Complete genome sequence of enterohemorrhagic Escherichia coli  
O157:H7 and genomic comparison with a laboratory strain K-12  
DNA Res. 8 (1), 11-22 (2001)  
21156231  
5 (bases 1 to 281530)  
Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H. and  
Hayashi, T.  
Direct Submission  
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome  
Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,  
Japan (E-mail: ken@gen-info.osaka-u.ac.jp,  
URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365,  
Fax: 81-6-6879-2047)  
genome project.

Location/Qualifiers  
1. 281530  
/organism="Escherichia coli O157:H7"  
/strain="O157:H7"  
/sub\_strain="RIMD 0509952"  
/db\_xref="taxon:83334"  
190..273  
/gene="Ecs0001"  
190..273  
/gene="Ecs0001"  
/note="similar to THRL\_ECOLI g11786182 percent similarity  
100 in 21 aa, but has 6 additional residues (Conserved in  
E.coli K-12)"  
/codon\_start=1  
/transl\_table=11  
/evidence=not experimental  
/product="thr operon leader peptide"  
/protein\_id="BAB33424.1"  
/db\_xref="GI:13359457"  
/translation="MKRISTITITITITITITITITITITGNGAG"  
354..2816  
/gene="Ecs0002"  
354..2816  
/gene="Ecs0002"  
/note="similar to THRA\_ECOLI g11786183 percent identity  
99 in 820 aa (Conserved in E.coli K-12)"  
/codon\_start=1  
/transl\_table=11  
/evidence=not experimental  
/product="aspartokinase I-homoserine dehydrogenase I"  
/protein\_id="BAB33425.1"  
/db\_xref="GI:13359458"  
/translation="MRVLFKFGTSSVANAERFLRVADILESNAEQGVATVLSAPAKIT  
NHLVAMIEKTSIGDAPNLSDAERIFAEILLTGLAAQGFPLAQLKTFVDFEFAQIK  
HVLHGISILGQCPDSINAALICRGKMSIATMAGLEARGHNTVIDPVKLLAVGHY  
LESTVDIAESTERIAASRIPADHVMYLMAGTAGNEKGLVYLRNGSDYSAVLAACL  
RADCCIEITWDVGVYTCDPQPRQVDFARLLKSMYSIQEAMLSYFGAKVLPRITPTIAQF  
QIPCLIKNTQPAFTGLTIGASRDEDELFPVKGISLNNMAMFSVSGPGMKGMVGAAR  
VFAMASRARSIVLLITSQSSYSISFCVPQSDCVRAERAMQEEFLKGLLEPLAV  
TERLAIISVVGDMRTLAGISAKFFAALARANINIVIAQGSSESSISVYVNDATT  
GVRYTHOMLENTDQVIEVFGVGVGGALLPOLKROOSLKNKHIDLRVCGVANSKA  
LITNVHGLNLENWQELAQAEPPNGLRLRLVKEYHLLNPVVDCTSSQAVADQYAD  
FLREGFHVVTNNKANTSSMDYTHLLRHAEEKRRKFLYDYNVAGLVPVLENQLNLLN  
AGDELMKFSGILSLSYIFGKLDEGMSFSEATTLAREMGYTEPPDRDLSGMVARK  
LLILARETGRELELADIEPVLPAEFNAEGDVAAFMANLSQDLDLFAARVAKRDEG  
KVLRVGNIDEGVCRVKAIEVDGNDPLFKVKNGENALAFYSHYQPLPLVLRGYGAG  
NDVTAAGVFADLLRLTSLNKLGV"  
2818..3750  
/gene="Ecs0003"  
2818..3750  
/gene="Ecs0003"  
/note="similar to THRB\_ECOLI g11786184 percent identity  
99 in 310 aa (Conserved in E.coli K-12)"

```

/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="homoserine kinase"
/protein_id="BAB33426.1"
/db_xref="GI:13359459"
/translation="MKVYAPASNMVSQFVLDGAATVPVDGALIGDVVTVESAEFF
SNNLGRFADKLPSEPRENIVVQWERFQELGKQIPVAMTLEKNMPPIGSLGSSACS
VVALMANHEGCKPLNDRLLALMGEALRGSRGTHYDNVAPCFPLGGWQMLTEENDI
ISOQVPGFDEWLVLAYPGIKVYSTAGARAILPAQYPRQDCIAHGRHLAGFIHACYSRQ
PELAALMKDVIAEYRERLLPGFRQARVAEIGAVASGISGSGPTLALCDKPDPA
QRVADWLGNYLQNOBGFVHICRLDTAGARVLEN"
3751..5037
/gene="ECs0004"
CDS
3751..5037
/gene="ECs0004"
/notes="similar to THRC_ECOLI g11786185 percent identity
99 in 428 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="threonine synthase"
/protein_id="BAB33427.1"
/db_xref="GI:13359460"
/translation="MKLYNLKDHNEQVSFAQVTOGLGRNQGLFFPHDLPEFSLTEID
EMLKDFVTRSAKILSAFTGDEIPOEILERVRAAFAPFAPVANVEDVGCLELPHGP
TLAFKDFGGRNAOMLTHTAGDKPVTILTATSGDTCAAAVAHAFYGLPNVKNVILYPRG
KISPLEKIFCTYLGNGIETVAIDGDEACQALYKQAFDDEELKVALGLNSANISLR
LQAQICFYEAVALQPEARNQLVVSPVSGNFGDLTAGLSKLGLPLVPRKFAANTNVN
DIVRFLDYGQSPKATQATLSNMDVSPQNNMVRVEELFRKKIWLKELGVAANDVE
TTQDMLKEALGELGYTSEPHAAVAYRALRDLNPGEGYGLFGLTAHPAKFKESVEAILGE
TLDLPELAERADLPILLSHNPADFAALRKLMMNHQ"
5251..5547
/gene="ECs0005"
CDS
5251..5547
/gene="ECs0005"
/notes="similar to B0005_ECOLI g11786186 percent identity
95 in 98 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB33428.1"
/db_xref="GI:13359461"
/translation="MKMQSIVLALSILVLPAMTAAEITLVPSVKLQIGDRDNRGY
YNDGGHWRDHGKWOHYEWRGNRHPGPPPPRHHKKAHHDHGHGGRKHR"
complement(5700..6476)
/gene="ECs0006"
CDS
complement(5700..6476)
/gene="ECs0006"
/notes="similar to YAAA_ECOLI g11786187 percent identity
99 in 258 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB33429.1"
/db_xref="GI:13359462"
/translation="MLILISPAKTLDYQSPLTTRYTLPELLDNSQOLIHAEKLTTPP
QISTLMRISDLAGNARFHDWQPDFTPENARQAILAFKGDVYITGLQAEITFSEDDFD
FAQOHLRLSGELVLRPLDLMQYPRLEGNARLENARKDLYQFWGDIITNKLNALA
AQGDNVVINSLEDFYKSVKPKLNAEITKPVFLDEKNGKFKIISFYAKKARGLMSRF
IITENRITRPEQUTGNSGEGYFFDEASSNGELVFKRYEQR"
complement(6546..7976)
/gene="ECs0007"
CDS
complement(6546..7976)
/gene="ECs0007"
/notes="similar to YAAJ_ECOLI g11786188 percent identity
99 in 476 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="putative inner membrane transport protein"

```

```

/protein_id="BAB33430.1"
/db_xref="GI:13359463"
/translation="MPDFFSFINSVLWGSVMYILFLGAGCWFTFRTGCVQFVRIROPFG
KSLKNSIHQPQGLTSEFQSLCTSLAARVSGNLAGVALAITAGGPGAVFWMWAAFGIG
QANAVARALSSFSFDPPLVTGIIILAVFALLAIIIRGLHGVARLMQGFVPLMAIITWILT
SLVICVINTQGLPHVTSIPESAFGWEAGGAAGYTLSQAITNGFQSRMFSNEAGMG
STPNAAAAASWPPHAAQGIQVIGIFIDTLVITASAMLLLAGNGTTPMPLEGIQ
LTKAMRVLMGSGWAEFTLVLTILFAESSIVANVYAENNLFFLRLLNPKAIWCLRTIC
TPATVIGGTLISLPLMWQLADIIMACMAITNLAILLLSPVHTIASDYLQRKLGVR
PVFDPLRYDFDQRLSRDAWDDVSQE"
8255..9208
/gene="ECs0008"
CDS
8255..9208
/gene="ECs0008"

alignment_scores:
  Quality: 1212.00      Length: 848
  Ratio: 2.200          Gaps: 15
  Percent Similarity: 64.976  Percent Identity: 33.844

alignment_block:
US-09-701-711-2 x AP002550 ..

Align seg 1/1 to: AP002550 from: 1 to: 281530

10 GlnValSerAlaMetThrMetAlaValMetMetValMetSer..... 23
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
201260 CGCATAATAAGATGCGGATGAAAAGTTGCTCATAGGCGTCTGCTGTT 201309
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
24 .....ThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIle 37
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
201310 TAGCAGCCCGCCGCTATACGCTGTAAGGGTTCTAGTAGGAAGATATTC 201359
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
37 hrIleThrGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeu 53
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
201360 ATTTCGAAGCGCTTCAGCGCTGCGCGTTGGTGGCGGCTCTCCTCAGTATG 201409
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
54 ProPheArgLeuGlyGlnValValSerGluAsnGlnLeuAlaAspGlyVa 70
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
201410 CCGTGCCGACAGCGGCACGGTTAATGATGAAGATATCAGTAATACCAT 201459
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
70 llyAlaLeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisG 87
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
201460 TCGCGCTCTGTTGCTACCGCAACTTTGAGGATGTTCCGCGTCTCTG 201509
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
87 lnGluGlyArgIleLeuTyrGlnValThrGluArgProLeuIleAlaGlu 103
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
201510 ATGTTGATACCTTCTTGGTTTCAGGTAAAGAACGTCGACCATTCGACG 201559
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
104 lIeAsnPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluG 120
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
201560 ATTACTTCTCCGGTAACAAATCGGTGAAGATGACATGCTGAAGCAAAA 201609
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
120 yLeuLysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrV 137
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
201610 CCTCGAGCTTCTGGTGGCGGTGGGGAATCCCTCGATCGCACCAACA 201659
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
137 alGlnMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyr 153
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
201660 TTGCCGATATCGAAGAGTCTGGAAGACTTCTACTACAGCGTCGGTAAA 201709
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
154 TyrAsnThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgVa 170
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
201710 TATAGCGCCACGGTTAAGCTGTCGTGACCCCGTGCCTGCGCAACCGTGT 201759
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
170 lLysLeuAspMetThrPheAlaGluGlyLysProAlaArgValValAspI 187
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
201760 TGACCTAAACTGGTGTTCAGGAAGGTGTGTGACGTCAAAATCCAGCAA 201809
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
187 leAsnIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspVal 203
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
201810 TTAACATTGTTGTAACCATGCTTTCACCCGACGACGATCTCTCTCAT 201859
: : : : : : : : : : : : : : : : : : : : : : : : : : : :

```







[illegible]

```

162  ACATCCGAGCACCCTCTGATCAGGATAACACGTTCAAACCGGACGACTTCA 1711
579  hrTyAsnAlaIleuGlyTrpAsnTyrSerSerLeuAspArgProVal 595
      |||||::||::||::||::||::||::||::||::||::||::||::||
1712  CGTCCAAC.....TATGGTTGGACCTATAACAAGCTTGACCGCTGGTTAC 1755
596  pheProThrGlnGly.....MetSerHisSerValAspLeuThrValGI 610
      |||||::||::||::||::||::||::||::||::||::||::||::||
1756  TTCGCCGACAGATGGTTACAGTGTCAACCTGACCGGTAAAGTGACCATTCC 1805
610  yPheGlyAspLysThrHisGlnLysValValTyArgGlyGlnGlyAsnIleTyA 627
      |||||::||::||::||::||::||::||::||::||::||::||::||
1806  TGGATCGGATACCAAGTACTACAAGTGACGTAGACACCGCGACTTATG 1855
627  tgpProPhe.....IleLysLysSerValLeuArgGlyTyrAlaLysLeu 641
      |||||::||::||::||::||::||::||::||::||::||::||::||
1856  TGCCGATCGATGACGATCACAAATGGGTGTCTGGCGCGTACCCCGCTGG 1905
642  GlyTyrGlyAsnAsnLeu.....ProPheTyrGluAsnPh 653
      |||||::||::||::||::||::||::||::||::||::||::||::||
1906  GGTATGATGGATGGTTTAGCGCGCAAGAGATGCGGTCTTACGAGAACTT 1955
653  etyAlaGlyGlyTyrGlySerValArgLysTyArgPlnSerSerLeu 670
      |||||::||::||::||::||::||::||::||::||::||::||::||
1956  CTATCCCGTGTGTTCCACGACCCGTCGCTTCCAGTCCCAATACCATTTG 2005
670  lyProArgSer.....GlnAla..... 675
      |||||::||::||::||::||::||::||::||::||::||::||::||
2006  GTTCCGAAGACAGTTTACTTCCCGCATGACGCCAGTAATATGATCGGGAC 2055
676  ....TyrLeuThrAlaArgGlyGlnGlnThrThrLeu..... 687
      |||||::||::||::||::||::||::||::||::||::||::||::||
2056  FATGATTACGAATGTGCGACTCAGAGCGCGCGAAACACCTGTGTAATC 2105
688  .GlyGluValValGlyGlyAsnAlaLeuAlaThrPheGlySerGluLeuI 704
      |||||::||::||::||::||::||::||::||::||::||::||::||
2106  GGATGATGCTGTAGCGGTAAACGGCATCGCGGTGTGCCAGCGCTCGAGTCA 2155
704  leLeuProLeuProPheLysGlyAsp...TrpIleAspGlnValArgPro 719
      |||||::||::||::||::||::||::||::||::||::||::||::||
2156  TCACCCCGCGCGGTATTATAGCATAGTATGCTAACTCGGTTCGTACT 2205
720  ValIlePheIleGluGlyGlyGlnValPheAspThrThrGlyMetAspLys 736
      |||||::||::||::||::||::||::||::||::||::||::||::||
2206  TCTCTTCTCTGGGATGTGGTACCGTGTGGGATACAACTGGGATCCAG 2255
736  scIlnThrIleAspLeuThrGlnPheLysAspProGlnAlaThrAlaGluG 753
      |||||::||::||::||::||::||::||::||::||::||::||::||
2256  CCAATATTCTGGATATCCGGACTATAGTGATCCA..... 2289
753  lnAsnAlaLysAlaAlaAsnArgProLeuLeuThrGlnAspLysGlnLeu 769
      |||||::||::||::||::||::||::||::||::||::||::||::||
2290  ....AGCAATATC 2298
770  ArgTyrSerAlaGlyValGlyAlaThrTrpTyrThrProIleGlyProLe 786
      |||||::||::||::||::||::||::||::||::||::||::||::||
2299  CGTATGCTTGGCGGTATCGCAATTCAATGGATGTCCCATTTGGGCGCGTT 2348
786  uSerIleSerTyAlaLysProLeuAsnLysLysGlnAsnAspGlnThrA 803
      |||||::||::||::||::||::||::||::||::||::||::||::||
2349  GGTGTTCTCTTACGCCCGCCGTTCAAAGAGTACGATGGAGACAAGCAG 2398
803  spThrValGlnPheGlnIleGlySerValPhe 813
      |||||::||::||::||::||::||::||::||::||::||::||::||
2399  AACACGTTCCAGTTTAACTCGGTAAACCTGG 2430
_name: gb_ba:AE004297
documentation_block:
US AE004297 10901 bp DNA linear
INM10N Vibrio cholerae chromosome I, section 205 of 251 of
chromosome.
AE004297
SESSION AE0043852

```



VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

AE004297.1 GI:9656810  
Vibrio cholerae.  
Vibrio cholerae  
Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
1 (bases 1 to 10901)  
Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.A., Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D., Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragol, I., Sellers, P., McDonald, L., Uitterback, T., Fleischmann, R.D., Nierman, W.C. and White, O.  
DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae  
Nature 406 (6795), 477-483 (2000)  
20406833  
2 (bases 1 to 10901)  
Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.A., Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D., Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragol, I., Sellers, P., McDonald, L., Uitterback, T., Fleischmann, R.D., Nierman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R., Mekalanos, J.J., Venter, J.C. and Fraser, C.M.  
Direct Submission  
Submitted (14-JUN-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA  
Location/Qualifiers  
1..10901  
/organism="Vibrio cholerae"  
/strain="N16961"  
/serotype="O1"  
/db\_xref="taxon:666"  
/chromosome="1"  
/note="biotype: El Tor"  
complement(69..1124)  
/gene="VC2250"  
complement(69..1124)  
/gene="VC2250"  
/note="similar to GB:X54797 SP:P21645 PID:1208950  
PID:41470 GB:U00096; identified by sequence similarity;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="UDP-3-O-3-hydroxymyristoyl glucosamine  
N-acyltransferase"  
/protein\_id="AAF95394.1"  
/db\_xref="GI:9656811"  
/translation="MTTLTAEATITGGELFGDPTALVSAPAMDKAQLGHVTFLSN  
PKYSHKLDCKATVMVKTSEELCPNSALVADPYVAFKVAQALDSTPSPAHGIAP  
SAVIAEDAKLGHNVISGANAVIESGVOLGDNVIGACGFCIGKQARLGDNTKLMANVTI  
YHKEIGSDCLIOSGTVIGADGFGYANERGIKIPOLGSRIGRDEIGACTIDRG  
ALDVTIEDNVIIQNIQAIHVHIGSALAGGVITAGSTRICKYICLIGASVINGH  
IEADGVITTMGMVMSRIEKGWTSIGIPLQTNKWRKTAARVHRIDDMHKLKALE  
KLIPQSDTVPQDNSQAE"  
complement(1134..1643)  
/gene="VC2251"  
complement(1134..1643)  
/gene="VC2251"  
/note="similar to PID:1935023; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="outer membrane protein OmpH"  
/protein\_id="AAF95395.1"  
/db\_xref="GI:9656812"  
/translation="MKNMKAASGLITLSSMNAEAAKIGYINTAQVFOALPQ  
REVLOKMQREFKDAELQADAKTKIEKLKRDGOLMGQDEKILRIEIGLDSK  
YKIAQALEQASARREAEKQKLFVITQDAVKVVAEREYDVIDLTSSMQYKPEHNL  
SEKVIKAIK"  
complement(1658..4069)  
/gene="VC2252"  
complement(1658..4069)

/gene="VC2252"  
/note="similar to GB:U00096 SP:P39170 PID:1552754  
PID:1786374; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="surface antigen"  
/protein\_id="AAF95396.1"  
/db\_xref="GI:9656813"  
/translation="MAMKQILLATLTSVANGAEKFWQDQIQIDGLQVALGALL  
KMPKVGSDVDSQVANIKAIISSGNFEDVKVLRDGNLTVQVQKERTPIASVSGN  
KAIKEQKQLEASSIRIQAALDRTLSNIEKLEDFIYSVGRKATVATKAVVPLPR  
NADIKFVTEGVSAKIQIINFIGNQVFSDELLSRFNLNDVAMNFIADKKYQV  
LAGDIEALRTYLDGRKIFQFQVSAISDPKGVITLNLNEGEPTVSVQVRGE  
LMGKEAFTSLIPPIGTYNGSAVTRLEESVKVLGSGVAYPOVRTIPRFDKQO  
VSLVHVHAGKRVYVRDTRFVGNNSRDEVLRNEMRQEGSWLSKDIETGKTRNL  
GFFTEVQTVRPGSDQVDLVISVREANGVNFVGYGTESVSGVQGLQQNRL  
GSGNRVGNVAMINDYKQNLLEYRDFYWNLDGSLGGKVFYQNFSEAGIVDYTNES  
YGTSLTFFPDELNRFEGIGYTHNKIGNLTPYLOVENFLAAQASNIDSGNLLTDD  
FDILNLSWTRNNLNSYFTAGNHQAFYKMTVPGSDAQYFKLQYDVQYFPLTKHEF  
TLLLRGLGNGYGTQDKNLPPFENFVAGGFTSLRFGSGSAGPKAVRYDSGS  
NGSDTATDSDVGGNAIALASVELIVPTPEASEARNOIRTSIFYDMASVNDTEFYR  
CKADYGNQYYDYSPTNRSYVALQWSPMGLPVFLSLAKPIKKYECDDDEPTFT  
IGRTF"  
complement(4118..5476)  
/gene="VC2253"  
complement(4118..5476)  
/gene="VC2253"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAF95397.1"  
/db\_xref="GI:9656814"  
/translation="MTDILWNFIATIALGILVAVHGRHFVWARRCGVKVEKFIGF  
GKSIWKRVDGTEYSISMLPGGVKMDGRVDVPAEQQAMAFDKSLMKRSIAVS  
AGPIFNFLFAIYAVLVMIPVAVKPVIGEVTPYSIAAQAGLEPMKIKAYSGVNT  
DWESYNGLIIGHIDDSMTITVSSAEGVGLNEIKTINLRDNFDPETESAMGALGFKP  
FTPEISNQLTNVSAQAGERAGLVDTLQINGQAVQVNAIQSHPNAPTAVM  
VERAQOVELTLPDSRELSOGKIVIGAGIAPKVAEPQNYRFEQVFGVESLGRAVE  
KSGQVIDLTVSMKKLLVGDYGLNLSGPISTAKGAGTADYGFYFGLGLAISNL  
RLNLVPLPLMDLGGHLLFFMIEAVIRRPVPEKVMQEMGRIGGAILIFSLMAVAFNDFT  
RL"  
complement(5473..6681)  
/gene="VC2254"  
complement(5473..6681)  
/gene="VC2254"  
/note="similar to SP:P45568 GB:U00096 PID:1552750  
PID:1786369; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="1-deoxy-D-xylulose 5-phosphate reductoisomerase"  
/protein\_id="AAF95398.1"  
/db\_xref="GI:9656815"  
/translation="MRKLTILGATSGISGASTLKVIQAQNPQFSIVALVAGVNAKMYQ  
LCQOWRPYAVATASAESEQLGLKNQAMATEVLYGEEAMQVLAALDDVDVTMAAIV  
GAAGLPTMAAVKAGKRVLLANKLVMSGOLFIDAVAQSAELMPDSEHNAIFQCL  
PTEICOTGRCDLSQHDHILLTSGSGPFRYSDLATLDSVTPEQAIAPHNWSMGPKI  
SVDSATMNMKGLYEYTEAKLWNTSREQLKVLHQPQSVIHSVVOYQDGSVIAQLGPEDM  
ATPISYAMAYPVRTAGVPALDFTRLQOLTTFMEVDFARYPCQLQIAMDACFLGQHATTS  
LNAANEVADAFALRRKIRFTDIALINDQVLSKVCATNTQLHCRDLESLELDTMARHF  
AHQVLEKROA"  
complement(6730..7572)  
/gene="VC2255"  
complement(6730..7572)  
/gene="VC2255"  
/note="similar to GB:I42023 SP:P44937 PID:1006039  
PID:1221024 PID:1205166; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="phosphatidate cytidyltransferase"  
/protein\_id="AAF95399.1"  
/db\_xref="GI:9656816"

/translation="MKQRIITAILAPLVILGILYLPFAWMIALAVVTLGLFWEWTQ  
FVNPQSRMLAMISFALLVGGISVALIDFQFPAISNMNTAHFIVLIGISLWLVSSGLAI  
TYPSRPLMHSTHRLFLGLFLLPFWSVLFRLADTVLSLDPLYGAKLVLFVCFVW  
AADSQAYFVKSLGKHKMAPVSPNKTIECLVGGIVTAMLVGYWVAECGIQPSMPV  
MLLIILITVVVISVIGDLVFSMFRVSGIKDSSNIIPGHGILDRISLTAAFPVFFALL  
YFLF"

gene complement(7590..8360)

/gene="VC2256"

complement(7590..8360)

/gene="VC2256"

/note="similar to GB:L42023 SP:P44938 PID:1005041

PID:1221025 PID:1205167; identified by sequence

similarity; putative"

/codon\_start=1

/transl\_table=11

/product="undecaprenyl diphosphate synthase"

/protein\_id="AAF95400.1"

/db\_xref="GI:9656817"

/translation="MFVFLMQNLSQDVLPHKIAIVMDGNRWAKAOKPRVFGHK  
NGVAARVKTISTSRERLOPKIAQAEQLASNTGMVNIAANYGGOWDILQATQALAE  
NNLRVIGDKSRESERLOPKIAQAEQLASNTGMVNIAANYGGOWDILQATQALAE  
KYTGELQSDIDEDVFKQHLTMDLDPDVLIRTSGECHSNFMQLWLAAYAEYFTF  
VFWPEDENCLIAVTFWVNRERRFGCTGEQVKAALMEN"

complement(8434..8991)

/gene="VC2257"

complement(8434..8991)

/gene="VC2257"

/note="similar to GB:L42023 SP:P44307 PID:1005819

PID:1220904 PID:1205056; identified by sequence

similarity; putative"

/codon\_start=1

/transl\_table=11

/product="ribosome recycling factor"

/protein\_id="AAF95401.1"

/db\_xref="GI:9656818"

/translation="MINEIKKDAQERMEKSVALKNGLSKIRTGRAHPSLLTGISVDY  
YGAPTLNQVANYIADARTLAITVDFRELTQKVERAILMSDILGNPMSAGTIIIRVPL  
PPTERRDLVKIVRGEABGGVAVRNIRRDNDLKLKDKKEISEDDERRAQBEI  
OKITDAKRRIDDLVLAKEKMEV"

complement(9045..9776)

/gene="VC2258"

complement(9045..9776)

/gene="VC2258"

/note="similar to GB:D26562 SP:P29464 GB:D13334 GB:X78809

PID:1208944; identified by sequence similarity; putative"

/codon\_start=1

/transl\_table=11

/product="uridylylate kinase"

/protein\_id="AAF95402.1"

/db\_xref="GI:9656819"

alignment\_scores:

Quality: 1203.00 Length: 833

Ratio: 2.168 Gaps: 15

Percent Similarity: 66.627 Percent Identity: 33.373

alignment\_block:

US-09-701-711-2 x AE004297/rev ..

Align seg 1/1 to reverse of: AE004297 from: 1 to: 10901

16 MetAlaValMetMetValMetSerThrHisAlaGlnAlaAlaAsp...Ph 31

4048 CTCGCGACACTGCTGCCACGAGCGTGTGGCGGAATGGTCGCGGAAATAAT 3999

31 eMetAlaAsnAspIleThrIleThrGlyLeuGlnArgValThrIleGlu 48

3998 TGTGTGACAGGATATCAAAATTGATGGCTGCAAGGGTCGCTTTGGGTG 3949

48 erLeuGlnSerValLeuProPheArgLeuGlyGlnValValSerGluAsn 64

3948 CGGCGTTGTTGAAATGCTGTTCGAGTCGGCGACAGTGTCTGATTCGCAA 3899

65 GlnLeuAlaAspGlyValYsalAlaLeuTyrAlaThrGlyAsnPheSerAs 81  
3898 GATGTCGCGAACATCATCAAGCGCTCTATTCTCAGGTAACCTTTTCAGGA 3849  
81 pValGlnValTyrHisGlnGluGlyArgIleIleTyrGlnValThrGluA 98  
3848 TGTCAAGTATTGCGTGTGCAATACCTTGATGTGTGAGGTTAAAGAGC 3799  
98 rgProLeuAlaGluIleAsnPheGluGlyAsnArgLeuIleProLys 114  
3798 GCCCACCATTGCAAGCGTGTCTTTTCGCCAACAAAGCGATCAAGAG 3749  
115 GluGlyLeuGlnGluGlyLeuLysAsnAlaGlyLeuAlaValGlyGlnPr 131  
3748 GAGCAACTCAAGCAAACTTGGAGCTTCGAGCATCCGAGTTGGCGAAGC 3699  
131 oLeuLysGlnAlaThrValGlnMetIleGluThrGluLeuThrAsnGlnT 148  
3698 CTTGTATCGAAGCAGCACATCAAGCAATATTGAAAAGGCTTAGAAGATTTT 3649  
148 yrIleSerGlnGlyTyrTyrAsnThrGluIleThrValLysGlnThrMet 164  
3648 ACTACAGCGTGGTAAATACAAAGCGACCGTGAAGCGGTGTAACCCCA 3599  
165 LeuAspGlyAsnArgValLysLeuAspMetThrPheAlaGluGlyLysPr 181  
3598 TTGCGCGCAACCGTGTGACCTTAAGTTTGTCTTACTGAGGCGGTTC 3549  
181 oAlaArgValValAspIleAsnIleIleGlyAsnGlnHisPheSerAspA 198  
3548 CGCTAAGATCCCAACAGATCAACTTTATCGGTAAACCAAGTGTTCGCGATG 3499  
198 laAspLeuIleAspValLeuAlaIleLysAspAsnLysIle.....Asn 212  
3498 AAGAGTTCCTGCTCAGCTTTTAACCTCAAGCTGTGATGCCATGGTGAAC 3449  
213 ProLeuSerLysAlaAspArgTyrThrGlnGluLysLeuValThrSerLe 229  
3448 TTCCCTTGGCGAT...GATAAGTACAAAGAACGAAAGTGTGGCGGGTGATAT 3402  
229 uGluAsnLeuArgAlaLysTyrLeuAsnAlaGlyPheValArgPheGluI 246  
3401 TGAACGCTGCGTACTTACTATCTGATCGCGGTATCTGAAGTTTCAGG 3352  
246 leLysAspAlaLysLeuAsnIleAsnGluAspLysAsnArgIlePheVal 262  
3351 TTGATAGCACACAGTGGCCATCTCTCCGATAAGAAAGGCGGTGATCATC 3302  
263 GluIleSerLeuHisGluGlyGlnTyrArgPheGlyGlnThrGlnPh 279  
3301 ACGCTGAACCTTGAATGAAGGTGAGCCGTCACGCGTAAGCAAGTGCAGTT 3252  
279 eLeuGlyAsnLeuThrTyrThrGlnAlaGluLeuGluAlaLeuLysP 296  
3251 CCGTGTGAGCTGATGGCAAGAGCGCGAGTTTACTTCTTATCCCAT 3202  
296 heLysAlaGluGlyPheSerGlnAlaMetLeuGluGlnThrThrAsn 312  
3201 TTGAAATCGCGGAAACCTATAACGGTTCTGCGGTGACTAGACTCGAAGAG 3152  
313 AsnIleSerThrLysPheGlyAspAspGlyTyrTyrTyrAlaGlnIleAr 329  
3151 AGTGTTAAGAAAGTCTGGGGAATCTGGCTACGCTATCCACAGTAGC 3102  
329 gProValThrArgIleAsnAspGluSerArgThrValAspValGluTyrT 346  
3101 CACCATTCCTGAATTTGATGATGAAAGCAGCAAGTGTCACTGGTGGTTC 3052  
346 yrIleAspProValHisProValTyrValArgIleAsnPheThrGly 362  
3051 ATGTCGAAAGCGGGTAAACGTTGTATGTCCGTGACATCGGTTTGTGCGT 3002  
363 AsnPheLysThrGlnAspGluValLeuArgArgGluMetArgGlnLeuGl 379



## AUTHORS

Simpson, A.J.G., Reinach, F.C., Arruda, P., Abreu, F.A., Acencio, M., Alvarenga, R., Alves, L.M.C., Araya, J.E., Baia, G.S., Baptista, C.S., Barros, M.H., Bonaccorsi, E.D., Bordin, S., Bove, J.M., Briones, M.R.S., Bueno, M.R.P., Camargo, A.A., Camargo, L.E.A., Carriaro, D.M., Carreir, H., Colauto, N.B., Colombo, C., Costa, F.F., Costa, M.C.R., Costa-Neto, C.M., Coutinho, L.L., Cristofani, M., Dias-Neto, E., Docena, C., El-Dorri, H., Facincani, A.P., Ferreira, A.J.S., Ferreira, V.C.A., Ferro, J.A., Fraga, J.S., Franca, S.C., Franco, M.C., Frohme, M., Furlan, L.R., Garnier, M., Goldman, G.H., Goldman, M.H.S., Gomes, S.L., Gruber, A., Ho, P.L., Hoheisel, J.D., Junqueira, M.L., Kemper, E.L., Kitajima, J.P., Krieger, J.E., Kuramae, E.E., Laigret, F., Lempers, M.R., Leite, L.C.C., Lemos, E.G.M., Lemos, M.V.F., Lopes, S.A., Lopes, C.R., Machado, J.A., Machado, M.A., Madeira, A.M.B.N., Madeira, H.M.F., Marino, C.L., Marques, M.V., Martins, E.A.L., Martins, E.M.F., Matsukuma, A.Y., Menck, C.F.M., Miracca, E.C., Miyaki, C.Y., Monteiro-Vitorello, C.B., Moon, D.H., Nagai, M.A., Nascimento, A.L.T.O., Netto, L.E.S., Nhani Jr., A., Nobrega, F.G., Nunes, L.R., Oliveira, M.A., de Oliveira, M.C., de Oliveira, R.C., Palmieri, D.A., Paris, A., Peixoto, B.R., Pereira, G.A.G., Pereira Jr., H.A., Pesquero, J.B., Quaggio, R.B., Roberto, P.G., Rodrigues, V., de M. Rosa, A.J., de Rosa Jr., V.E., de Sa, R.G., Santelli, R.V., Sawasaki, H.E., da Silva, A.C.R., da Silva, F.R., da Silva, A.M., Silva Jr., W.A., da Silveira, J.F., Silvestri, M.L.Z., Siqueira, W.J., de Souza, A.A., de Souza, A.P., Terenzi, M.F., Truffi, D., Tsai, S.M., Tsubako, M.H., Vallada, H., Van Sluys, M.A., Verjovski-Almeida, S., Vettore, A.L., Zago, M.A., Zatz, M., Meidanis, J. and Setubal, J.C.

## Direct Submission

Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP 13083-970, Brazil

## FEATURES

Location/Qualifiers  
1..11711  
/organism="Xylella fastidiosa 9a5c"  
/db\_xref="taxon:160492"  
/clone="9a5c"  
complement(84)..1424  
/gene="XF1037"  
complement(84)..1424  
/gene="XF1037"  
/note="similar to SPIP28183 (percent identity: 65 %/query alignment coverage: 100.0 %/subject alignment coverage: 96.3 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"  
/codon\_start=1  
/transl\_table=11  
/product="adenosylhomocysteinase"  
/protein\_id="AAF83847.1"  
/db\_xref="GI:9105979"  
/translation="MPGLMSIRRYASKQPLKGVYVTCGSLHMTQTAVLIELTKDIGA DWRWSCNIFSTQDHAAAIATSGTPFWAKGETLEEYDCTLQALFTTISDGLTGP ELIVDGDGDTLLHKGYELNGSTWDEPSDLEEQVIRKLLRIATIRPGYWRVY NDWKGVSRETTGVRLYQIAATGRLVPAINDSVTKSKFDNLGCRSLADGLKR AMDVMAKLVACGCGYDGVGKSAHSLRAYCARVITFDIPICALQAAEGFEVTTVE DTLGADIYVTTGNKQVIRIEHTAMKQDVIVCNIGHFNEIQVDALNALAGVQKIN IKPQDKFILPNGNTLFLIAEGRNLVNCATGHPFSFVMSNSFANQTLAQIDWQNDV YEKNYRLPKLDEEVARLHLEKIGVKLTTLTANQAAAYLISVEGPKPEHYR"  
2470..3981  
/gene="XF1038"  
2470..3981  
/gene="XF1038"  
/note="similar to GI12808752 (percent identity: 23 %/query alignment coverage: 104.4 %/subject alignment coverage: 98.7 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"  
/codon\_start=1  
/transl\_table=11  
/product="peptide synthase"  
/protein\_id="AAF83848.1"  
/db\_xref="GI:9105980"  
/translation="MVHCVCVLYSKRCSECTVHCYRRVAGLVLPVCMQTSVCLHGD DLLDCLYAMSLVAPSSSVASLPMGVGSLHRPLVWVGKPIITGLFLAORYGLAAOL PSGRVYNLCEDRFLFLAFYACALRGQVSLPSSRAPGVVAEIIYAAPDSYCLGDRE ILSLESLRYWHLLPEIFELDGAIPSLSDAVVAIGFTSGTGVTNSLNATKVTSTSE

gene  
CDS

QDWAFHSLMGSOVTVVATVPPOHMYGMEFVLLPLLPVAVVAVVAGRPFFPEDVAQAL AEYAPRLVTTPVHLRVQSLRVPPEVAGIVSATAPLTHDLAVAAEACFGCELREY FGSTETCFVARTVYEQMTRDGVSEIENADVTWYARHLPAAPMAIDVVDYEDG RFVILGQADLLEIAGKRASIGELNRCLLAIPGVVDGALYQMPLEADQTVGRIAVVV APTLHEVHIQVIRKIDPLFLPRLRRKVDALPNNETGKLPREALALLADVC"  
4249..4413  
/gene="XF1039"  
4249..4413  
/gene="XF1039"  
/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF83849.1"  
/db\_xref="GI:9105981"  
/translation="MLVNTSVCNVSWLVFTQHCLLGLVAEMRCLLIALLAWRWILVF SMGCKYRGIV"  
4430..4591  
/gene="XF1040"  
4430..4591  
/gene="XF1040"  
/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF83850.1"  
/db\_xref="GI:9105982"  
/translation="MAFTGCNAGCVGTFNFCVAGVQLGRGLMVCVIEWSVLGRSGFY HVSSSGVWS"  
complement(4780)..5484  
/gene="XF1041"  
complement(4780)..5484  
/gene="XF1041"  
/note="similar to SPIP10442 (percent identity: 55 %/query alignment coverage: 81.6 %/subject alignment coverage: 96.5 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"  
/codon\_start=1  
/transl\_table=11  
/product="ribonuclease HII"  
/protein\_id="AAF83851.1"  
/db\_xref="GI:9105983"  
/translation="MVSQFDTHLPSSTNVALVAGDEAGRGLAGPVPVVAAVFDPSQ PRINGLNDKSLSPACRERLYAHIVERALAYKVYMDISTQIDTINIQTAMGLMRLAV QYVAHVAKSARIDGNRLPKPCPAEALVGGDARDRTIMAAASILAKVTRDRHMYELHL EGYHYGDKHKGYGTPAHLAALAEHGPCLEHRRSFAPVRKMLTLEAIHARQSTHQHNE NSPTKAAFNMLLEROD"  
complement(5541)..6698  
/gene="XF1042"  
complement(5541)..6698  
/gene="XF1042"  
/note="similar to SPIP10441 (percent identity: 45 %/query alignment coverage: 97.4 %/subject alignment coverage: 98.2 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder/Start codon shift: -126"  
/codon\_start=1  
/transl\_table=11  
/product="lipid A disaccharide synthase"  
/protein\_id="AAF83852.1"  
/db\_xref="GI:9105984"  
/translation="MIQAPRIALITAGEASGDHLGAGLITQOLRLRLPTAEFVIGIGDMM RSACQWTFDTSELAVMGLTEVLRHLRLKIRREFCKRALAWHPDVLIGIDAPDFNL TVERWKORHIRTIVYVPSIWAWEKRAKIGASVDRLVCLFPMPEPTIYARYGIDAR FVGHPMADVEIPQTDRTARTALGLPLVPLVLPGRSHSEISOLGPTFLAAGOLDS BHLGLHVPIPAANTQCKPPLAEQLSRSTLPVMSHLSDNSARTAMLAADVVLVASGT ATLEAMLKLRPMVAIVAPLTIYRIVTKLKLKLNRFALPNILAGEDLAPELLIQDCT APALCAALLDFKPKQKVAATLRYLQTLQLRNNASTRAAEAITEILLPQR"  
complement(6775)..7566  
/gene="XF1043"  
complement(6775)..7566  
/gene="XF1043"

gene  
CDS



```
|||||..... |||||||..... |||||||..... |||||||..... |||||||.....
10249 GTGTCGTAGATGCTTCAGCGAGAGATACCAAGATCGCTATTGCTTCTC 10200
478 yMetThrAsnProTyrPheThrValAsnGlyValSerGlnSerLeuSerG 495
: |||||||..... |||||||..... |||||||..... |||||||..... |||||||.....
10199 CTACACCAACCGCTTTTCACCGCAATGGGCTCTCAGTACAGTATAACT 10150
495 lyTyrTyrArgLysThrLysTyrAspAsnLysAsnLysAsnLysAsnLysVal 511
|||||..... |||||||..... |||||||..... |||||||..... |||||||.....
10149 TGGCCTACCAAAATAGACTACTCAGACTTCAACGCTGCGCAATCAAC 10100
512 LeuAspSerTyrGlySerLeuSerTyrGlyTyrProIleAspGluAs 528
: |||||||..... |||||||..... |||||||..... |||||||..... |||||||.....
10099 AGCAAGCGTATGTCAGGACAAACCATCTTTGGAATCCCGATTACAGAA 10050
528 nGlnArgIleSerPheGlyLeuAsnAlaAspAsnThrLysLeuHisGlyG 545
|||||..... |||||||..... |||||||..... |||||||..... |||||||.....
10049 CGACACTGCTCTTGGGTGTCATGGCGGAGATAGCAACCAATC.....A 10006
545 lyArgPheMetGlyIleSerAsnValLysGlnLeuMetAlaAspGlyGly 561
|||||..... |||||||..... |||||||..... |||||||..... |||||||.....
10005 CGACCTTCCAGGT...TCCAGCGCAAAAGCAATATTATCGAC..... 9968
562 LysIleGlnValAspAsnAsnGlyIleProAspPheLysHisAspTyrTh 578
: |||||||..... |||||||..... |||||||..... |||||||..... |||||||.....
9967 .....TACATCGACGCGAGTCGA.....CAGCGTACCTTCGG 9936
578 rThrTyrAsnAlaIleLeuGlyTyrAsnTyrSerSerLeuAspArgProV 595
9935 CGCTTGGCGCACCGAGCTGGGCTGGCGAGTCGACACTCGCAACGACTACT 9886
595 alPheProThrGlnGlyMetSerHisSerValAspLeuThrValGlyPhe 611
: |||||||..... |||||||..... |||||||..... |||||||..... |||||||.....
9885 TCATGCCAAATCTTGTTATGATCAGCGCATTCGGCGCAGAGTACACATA 9836
612 GlyAspLysThr.....HisGlnLysValValTyrGlnGlyAsnIleTyr 626
9835 CGCGGCTCAACGATCAAGTACTACAAAATCAATACCAGATCTCGAAATA 9786
626 rArgProPheIleLysSerValLeuArgGlyTyrAlaLysLeuGlyT 643
9785 CTGGCCCAATTATCCCGCGCTAGTACTCAACACGCGCTTGAAGTAGGT 9736
643 yTGlyAsnAsn..... 646
9735 ACGGAGACGACTACGSAAGAGTCACTCGGATCTGCTCGCAGCGGACT 9686
647 .....LeuProPheTyrGluAsnPheTyrAlaGlyG 657
9685 GTGGCCACTGCTTCGGGATTCGCTTTTCGAAATTTCTACGCGCGGG 9636
657 yTyrGlySerValArgGlyTyrAspGlnSerSerLeuGlyProArgSerG 674
9635 CACCACTCGTACGAGGCTTCAGAGATAACACCTAGGACCGCGCTCGG 9586
674 lnaLysTyrLeuThrAlaArgArgGlyGlnGlnThrThrLeuGlyGluVal 690
9585 AA.....GTGACCGCACTGTACAATCAA.....GGCCAAACCA 9554
691 valGlyGlyAsnAlaLeuAlaThrPheGlySerGluLeuIleLeuProLe 707
9553 TTGGGTGGTTCCTTCAAGACTGTGGATCGACTGAAATGACTTCCCAA 9504
707 uProPheLysGlyAspTrpIleAspGlnValArgProValIlePheIleG 724
9503 GCTTTTCGACAGC.....CCATCGCGACGCTATCTCAGCTTTCCTTG 9463
724 luGlyGlyGlnValPheAspThrThrGlyMetAspLysGlnThrIleAsp 740
9462 ATTTCCGCAACGCTCTTCAAT..... 9443
741 LeuThrGlnPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLysAl 757
```

```
9442 .....GGCGTAAATAATAATTTCAAGGC 9423
757 aAlaAsnArgProLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaG 774
9422 A.....AACGAATTACGTGCTTCAGCG 9400
774 lyValGlyAlaThrTrpTyrThrProIleGlyProLeuSerIleSerTyr 790
9399 CGGTAGCGCTGCTATGCGGTGCCCCCATTTGGTCCGATCTCGATCAGCTAT 9350
791 AlalysProLeuAsnLysLysGlnAsnAspGlnThrAspThrValGlnPh 807
9349 GCATTTCTATAAAGAGGAAACACGACGAATCGACGCGACTGCAATT 9300
807 eGlnIleGlySerValPhe 813
9299 CACCTTCGCGCGCCCAATTC 9281
seq_name: gb_pat:AX155443
seq_documentation_block:
LOCUS AX155443 2379 bp DNA linear PAT 22-JUN-2001
DEFINITION Sequence 6 from Patent WO0138350.
ACCESSION AX155443
VERSION AX155443.1 GI:14536771
KEYWORDS
SOURCE
ORGANISM
Neisseria meningitidis.
Bacterium; Proteobacteria; beta subdivision; Neisseriaceae;
Neisseria. 1 (bases 1 to 2379)
REFERENCE
AUTHORS Giuliani,M.M., Pizsa,M., Rappuoli,R. and Holst,J.
TITLE 85kda neisserial antigen
JOURNAL Patent: WO 0138350-A 6 31-MAY-2001;
Chiron Spa (IT) ; Statens Institutt for Folkehelse (NO)
FEATURES
source
1..2379
/db_name="Neisseria meningitidis"
BASE COUNT 636 a 755 c 596 g 392 t
ORIGIN
alignment_scores:
Quality: 1123.50 Length: 837
Ratio: 2.073 Gaps: 19
Percent Similarity: 64.755 Percent Identity: 32.616
alignment_block:
US-09-701-711-2 x AX155443 ..
Align seg 1/1 to: AX155443 from: 1 to: 2379
10 GlnValSerAlaMetThrMetAlaValMetMetValMetSerThrHisAl 26
4 AAACCTGAACAGATTGCTCCGCACTGATGATGTTGGGCATATCGCTTT 53
26 aGlnAlaAlaAspPheMetAlaAsnAspIleThrIleThrGlyLeuGlnA 43
54 GGCATTTGCGGACTTCCACCATCCAGACATCCGTCGAAAGGCTTCGAGC 103
43 rgValThrIleGluSerLeuGlnSerValLeuProPheArgLeuGlyGln 59
104 GTACCGAGCGGAGCAGCGATTCCTCACTACCTGCCCCCAAGTCGGCGAC 153
60 ValValSerGluAsnGlnLeuAlaAspGlyValLysAlaLeuTyrAlaTh 76
154 ACCTACACGACACACACGCGAGTCCATCATCAAAAGCCTGTACGCCAC 203
76 rGlyAsnPheSerAspValGlnValTyrHisGlnGluGlyArgIleIleT 93
204 CGGTTTCTTTGACGAGTACGAGTCAAACTCGCGAGCGGCGAGCTTCTGC 253
```





1979	CGGCTACGAAGCCGGCAGCTCGGCCGAAAA.....	2010
679	AlaargArgelyGlnThrLeuGlygluValVal.....GlyGI	693
2011	.....GTGTATGCAATAACGGCGAATAATACGCTACGGCGG	2048
693	yAsnAlaLeuAlaThrPheGlySerGluLeuIleLeuProLeuProPheL	710
2049	CAACAAAAGCCAAACGCTCTCCGCGAGCTGCTTCCCGATCCCGGGTG	2098
710	ysGlyAspTrpIleAspGlnValArgProValIlePheIleGluGlyGly	726
2099	CGAAGAC...GCACGACCCTCGCCTGAGCTGTTCGCGAGCAGGC	2145
727	GlnValPheAspThrThrGlyMetAspLysGlnThrIleAspLeuThrGl	743
2146	AGCGTGTGG.....GACGCGAGAACCATTATACCGCGCGCGA	2180
743	nPhelyAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAsnA	760
2181	AACGGTAACAACAATTCGGTTACTCGGAACGCGCATAAATCC....	2226
760	rgProLeuLeuThrGlnAspLysGlnLeuArgTyrsrAlaGlyValGly	776
2227	.....ACCTTACCAACAAATTCGCTATTCCGCGCGCGCGCG	2265
777	AlaThrTrpTyThrProIleGlyProLeuSerIleSerTyrrAlalysPr	793
2266	GTTACCTGGCTCTCGCCTTTGGTCCGATGAAATTCAGCTACGCTACCC	2315
793	OleuAsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGlnIleG	810
2316	GCTGAAGAAAAACCGGAGACGAAATCCAACGCTTCCAATTCAGCTCG	2365
810	lySerValPhe 813	
2366	GCACGACGTT C 2376	
seq_name:	gb_pat:AX202493	
seq_documentation_block:		
LOCUS	AX202493	2379 bp DNA linear PAT 30-AUG-2001
DEFINITION	Sequence 6 from Patent WO0152885.	
ACCESSION	AX202493	
VERSION	AX202493.1	GI:15392216
KEYWORDS	Neisseria gonorrhoeae.	
SOURCE	Neisseria gonorrhoeae.	
ORGANISM	Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;	
REFERENCE	1 (bases 1 to 2379)	
AUTHORS	Pizza,M.C., Rappuoli,R.C. and Giuliani,M.C.	
TITLE	Outer membrane vesicle (omv) vaccine comprising N. meningitidis serogroup b outer membrane proteins	
JOURNAL	Patent: WO 0152885-A 6 26-JUL-2001;	
Chiron Spa (IT)		
FEATURES	Location/Qualifiers	
source	1..2379	
/organism=	"Neisseria gonorrhoeae"	
/db_xref=	"taxon:485"	
BASE COUNT	636 a 755 c 596 g 392 t	
ORIGIN		
alignment_scores:		
Quality:	1123.50	Length: 837
Ratio:	2.073	Gaps: 19
Percent Similarity:	64.755	Percent Identity: 32.616
alignment_block:		
US-09-701-711-2 x AX202493		
Align seg 1/1 to: AX202493	from: 1 to: 2379	



```

308 GluGlnThrThrAsnAsnIleSerThrLysPheGlyAspGlyTyrTy 324
   :::::::::::::::::::::|||||
904 ACCCGCGTTTGGGTGAGATTGAGAACCGCATGGCTCGCGAGGCTACGC 953
   :::::::::::::::::::::|||||
324 rTyrAlaGlnIleArgProValThrArgIleAsnAspGluSerArgThrV 341
   :::::::::::::::::::::|||||
954 ATACAGCGAAATCAGCGTACAGCGCTGCGCAACCGCGGCAACCAACCG 1003
   :::::::::::::::::::::|||||
341 alAspValGluTyrTyrIleAspProValHisProValTyrValArgArg 357
   :::::::::::::::::::::|||||
1004 TCGATTTCGTCCTGCATCGACACCGCGCGGAAATCTACGTCAAGCAA 1053
   :::::::::::::::::::::|||||
358 IleAsnPheThrGlyAsnPheLysThrGlnAspGluValLeuArgArgG1 374
   :::::::::::::::::::::|||||
1054 ATCCACATCACCAGCAACAACAACCCGCGAGCGAGTGTGCGCGCGA 1103
   :::::::::::::::::::::|||||
374 uMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnLeuS 391
   :::::::::::::::::::::|||||
1104 ATTGCGCAATGGAAATCCCGCGCTTACGACACCTTCCAAAGCTGCAACGCT 1153
   :::::::::::::::::::::|||||
391 erArgAlaArgLeuMetArgThrGlyPhePheLysHisValThrValAsp 407
   :::::::::::::::::::::|||||
1154 CCAAAGACGCGTCCGACTTTGGCTACTTCGACACGCTACAGTTTGAT 1203
   :::::::::::::::::::::|||||
408 ThrArgProValProAsnSerProAspGlnValAspValAsnPheValVa 424
   :::::::::::::::::::::|||||
1204 GCGTCCCGCTTGGGTGATGTCGCGCGGTATCGCAGGACACCTGTTCGGT 1253
   :::::::::::::::::::::|||||
424 lGluGlnProSerGlySerThrIleAlaAlaGlyTyrSerGlnS 441
   :::::::::::::::::::::|||||
1254 GACGGAACGCTCCACCGCGCTCGACTTGGCGCGGCTGGGTTCAGG 1303
   :::::::::::::::::::::|||||
441 erGlyGlyValThrPheGlnPheAspValSerGlnAsnAsnPheMetGly 457
   :::::::::::::::::::::|||||
1304 ATACCGCTTGGTTCATGTCGCGCGGTATCGCAGGACACCTGTTCGGT 1353
   :::::::::::::::::::::|||||
458 ThrGlyLysHisValAsnAlaSerPheSerArgSerGluThrArgGluVa 474
   :::::::::::::::::::::|||||
1354 ACGGCAAGTCGCGCGCTCGCGCGCTCGCAAGCAAAACACGCTCAA 1403
   :::::::::::::::::::::|||||
474 lTyrSerLeuGlyMetThrAsnProTyrPheThrValAsnGlyValSerG 491
   :::::::::::::::::::::|||||
1404 CGGCTCGCTCGCTTACCGACCGGTACTTCAGCGCAGCGGGTCCAGC. 1452
   :::::::::::::::::::::|||||
491 lNserLeuSerGlyTyr.....TyrArgLysThr.....LysTyr 502
   :::::::::::::::::::::|||||
1453 .....CTGGGCTAGATATTACGAAAGACCTTCGACCGCGGCAAA 1494
   :::::::::::::::::::::|||||
503 AspAsnLysAsnIleSerAsnTyrValLeuAspSerTyrGlyGlySerIle 519
   :::::::::::::::::::::|||||
1495 GCATCGACACGCGTCAACAATATAAACCCACCACCGCGCGCGCGCT 1544
   :::::::::::::::::::::|||||
519 uSerTyrGlyTyrProIleAspGluAsnGlnArgIleSerPheGlyLeuA 536
   :::::::::::::::::::::|||||
1545 AAGATGGGTATCCCGTTTACCGAATACGACCGCTCAATTTCCGGCTGG 1594
   :::::::::::::::::::::|||||
536 snAlaAspAsnThrLysLeuHisGly..... 544
   :::::::::::::::::::::|||||
1595 CGCGGGAACCTTCAGCGTCAACACCTACAAACAAAGCACCCAAACGCTAT 1644
   :::::::::::::::::::::|||||
545 .....GlyArgPheMetGlyIleSerAsnVally 554
   :::::::::::::::::::::|||||
1645 GCCGACTTATCAGGAATACGCGCAAAACCGACGCG..... 1680
   :::::::::::::::::::::|||||
554 sGlnLeuMetAlaAspGlyGlyLysIleGlnValAspAsnAsnGlyIleP 571
   :::::::::::::::::::::|||||
1681 .....GCAGACGCG..... 1689
   :::::::::::::::::::::|||||
571 roAspPheLysHisAspTyrThrThrTyrAsnAlaIleLeuGlyTyrAsn 587
   :::::::::::::::::::::|||||
1690 ..AGCTTCAAAGG.....CTGCTGTACAAAGCAGCGCTCGGCTGGGG 1731
   :::::::::::::::::::::|||||
588 TyrSerSerLeuAspArgProValPheProThrGlnGlyMetSerHisSe 604

```

```

1732 CGCAACAAGACCGACAGCGCGTCTATGCCGACGCGGCTACCTGACCGG 1781
   :::::::::::::::::::::|||||
604 rValAspLeuThrValGlyPhe...GlyAspLys...ThrHisGlnLysV 619
   :::::::::::::::::::::|||||
1782 CGTAATACGCCAAATCCCTGCCCGCGCAGCAAACTGCAATACTACTACCC 1831
   :::::::::::::::::::::|||||
619 alValTyrGlnGlyAsnIleTyrArgProPheIleLysLysSerValLeu 635
   :::::::::::::::::::::|||||
1832 CCACCCACACCAACCTGGTTCCTCCCTTAACCAAAACCTTCACCGGTG 1881
   :::::::::::::::::::::|||||
636 Arg.....GlyTyrAlaLysLeuGlyTyrGly.....As 645
   :::::::::::::::::::::|||||
1882 ATGCTCGCGCGCAAGTCGCGATTGCGGCG...GGCTACGCGCAGAACAA 1928
   :::::::::::::::::::::|||||
645 nAsnLeuProPheTyrGluAsnPheTyrAlaGlyGlyTyrGlySerValA 662
   :::::::::::::::::::::|||||
1929 AGAAATCCCTTCTTTGAAACCTTCTACGGCGCGGCTGGGTTCGGTGC 1978
   :::::::::::::::::::::|||||
662 rgGlyTyrAspGlnSerSerLeuGlyProArgSerGlnAlaTyrLeuThr 678
   :::::::::::::::::::::|||||
1979 GCGGCTACGAAGCGGCGACGCTCGGCCGAAA..... 2010
   :::::::::::::::::::::|||||
679 AlaArgArgGlyGlnGlnThrThrLeuGlyGluValVal.....GlyG1 693
   :::::::::::::::::::::|||||
2011 .....GTGTATGACGAATACGCGCAAAATCAGCTACGCGCG 2048
   :::::::::::::::::::::|||||
693 yAsnAlaLeuAlaThrPheGlySerGluLeuIleLeuProLeuProPheL 710
   :::::::::::::::::::::|||||
2049 CACAAAAAAGCCAACTCTCCGCGGAGTCTCTCCGATGCCGCGGTG 2098
   :::::::::::::::::::::|||||
710 ysGlyAspTrpIleAspGlnValArgProValIlePheIleGluGlyGly 726
   :::::::::::::::::::::|||||
2099 CGAAAGAC...GCACGACCGCTCGCGCTGAGCCCTGTTGCGGACGCGAGC 2145
   :::::::::::::::::::::|||||
727 GlnValPheAspThrThrGlyMetAspLysGlnThrIleAspLeuThrGl 743
   :::::::::::::::::::::|||||
2146 AGCGTGTGG.....GACGCGAAGAACCTATACCGCGCGCGA 2180
   :::::::::::::::::::::|||||
743 nPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnA 760
   :::::::::::::::::::::|||||
2181 AACGGTAACAACAATCGGTTCCTCGAAACCGCGCATAAATCC.... 2226
   :::::::::::::::::::::|||||
760 rgProLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGly 776
   :::::::::::::::::::::|||||
2227 .....ACCTTTACCAACGAATTCGCTATTCCGCGCGCGCGCG 2265
   :::::::::::::::::::::|||||
777 AlaThrTrpTyrThrProIleGlyProLeuSerIleSerTyrAlaLysPr 793
   :::::::::::::::::::::|||||
2266 GTTACCTGCGCTCGCGCTTGGTTCGATGAAATTCAGCTACGCTACCC 2315
   :::::::::::::::::::::|||||
793 oLeuAsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGlnIleG 810
   :::::::::::::::::::::|||||
2316 GCTGAAGAAAAACCGGAAGACGAATCCAACGCTTCCAATTCAGCTCG 2365
   :::::::::::::::::::::|||||
810 lySerValPhe 813
   :::::::::::::::::::::|||||
2366 GCACGAGCTTC 2376
   :::::::::::::::::::::|||||
seq_name: gb_pat:AX155439
seq_documentation_block:
LOCUS AX155439 2394 bp DNA linear PAT 23-JUN-2001
DEFINITION Sequence 2 from Patent WO0138350.
ACCESSION AX155439
VERSION AX155439.1 GI:14536770
KEYWORDS
SOURCE
ORGANISM
Neisseria meningitidis.
Neisseria meningitidis
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
Neisseria.
REFERENCE
1 (bases 1 to 2394)
AUTHORS
Giuliani,M.M., Pizza,M., Rappuoli,R. and Holst,J.

```

TITLE	85kda neisserial antigen
JOURNAL	Patent: WO 0138350-A 2 31-MAY-2001; Chiron SA (fr) ; Statens Institut for Folkehelsetse (NO)
FEATURES	Location/Qualifiers
source	1. .2394
	/organism="Neisseria meningitidis" /db_xref="taxon:487"
BASE COUNT	657 a 743 c 583 g 411 t
ORIGIN	

alignment_scores:		
Quality:	1115.00	Length: 819
Ratio:	2.042	Gaps: 14
Percent Similarity:	66.667	Percent Identity: 32.479

alignment\_block:

US-09-701-711-2 x AX155439

Align seq 1/1 to: AX155439 from: 1 to: 2394

10 GlnValSerAlaMetThrMetalValMetMetValMetSerThrHisal 26  
::: :::: ::::: |||:::||||||::: :::  
4 AAACGAAACAGATTGCTTCGCACATGATGATGTGGCATCATCGCCTT 53

26 aGlnLaAlaAspPheMetAlaAsnAspIleThrIleThrGlyLeuGlnA 43  
||||||| ::::: ||||| ::: ||||||||  
54 GGCACTTGCCGACTTCACCATCCGAAGACATCCGGCTCGAAGGCTTCGAC 103

43 rgValThrIleGluSerLeuGlnSerValLeuProPheArgLeuGlyGln 59  
||: ::::: ::::: ||||| ::::: |||||  
104 GTACCGCGCGAGTACCGTATTCAACTACCTGCCCGCTCAAGTCGCGGAC 153

60 valValSerGluAsnGlnLeuAlaAspGlyValLysAlaLeuTyrlath 76  
::: ::::: ::::: |||:::|||||||  
154 ACCTCAACGACACACACGCGGACGCCATCATCAAAAGCCTGTACGCCAC 203

76 rGlyAsnPheSerAspValGlnValTyrlHisGlnGluGlyArgIlelrm 93  
||| ||||| ::|||::: ||| ::|||::: |||  
204 CGTTTCTTGACGAGGTACCGCTCGGAACCTCGGACGGCACGCTCTGC 253

93 yrGlnValThrGluArgProLeuIleAlaGluIleAsnPheGluYAsn 109  
||| ||||| ||||| ::|||::: |||  
254 TGACCGTTTATCAACGCCCCACCATTGCGCTCGCTCAACATCACCGCGCA 303

110 ArgLeulleProGlySgluGlyLeuGlnGluLeuLysAsnlaGlyLe 126  
::: ::::: ::::: |||:::|||||||  
304 AAAATGCTCAAAAACGACGCCATTAAAGAAAAACCTCGAATCGTTCGGGCT 353

126 uAlaValGlyGlnProLeuLysGlnAlaThrValGlnMeIleGluThrC 143  
||| ::||| ::|||::: |||||::: |||::: :::  
354 GGGCAGTCGCACAATCTTTATCAGGCGACACTCAATCAGCGACTCGCG 403

143 luLeuThrAsnGlnTyrlleSerGlnGlyTyTrYAsnThrGluIleThr 159  
||| ::||| ::|||::: ||| ||| ::|||  
404 GCCTGAAAGAAGAAATACCTCGGCGCGCAAACTCAATATCCAATCACG 453

160 valLysGlnThrMetLeuAspGlyAsnArgValLysLeuAspMethrPh 176  
||| ||| ||| ||||| |||  
454 CCCAAAGTAACCAAATCTCGCGCAACCGCGCTCGACATCGACATCAGAT 503

176 eAlaGluGlyLysProAlaArgValAlaSpileAsnIleleGlyAsnC 193  
: ||||| ||| ::|||::: |||||::: |||  
504 TGACGAGGCAAAATCCGCCAAATACCGACATCGAATTTGAAGGCAACC 553

193 lInHisPheSerAspAlaAspLeuIleAspValLeuAlaIleLysAspAsn 209  
|| ||||| |||||  
554 AAGTCTATTCCGACCGCAAACTGATCGCGCAAAATGTCCTCGACCGAAGC 603

210 LysIle...AsnPProLeuSerLysAlaAspargTyrrThrGlnGlulysLe 225  
||| ::: |||::: |||::: |||::: |||  
604 GGCATTGTGCACATGGCTACGACGAAGCAACCAATTTCAACGACGACGAAT 653

[illegible]





```

2194 .....GGCGGCGGGTTCAAACATTTACGCGCGCGCAATACCCA 2234
|||||
761 oLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaT 778
|||||
2235 TAAATCCACCTTTACCAACGAATTCGCGTAATCCGCGCGCGCGGTGTA 2284
|||||
778 hrTptTyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeu 794
|||||
2285 CCGTGCCTCGCCTTTAGGCCGATGAATTCAGCTACGCTACCGCTG 2334
|||||
795 AsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGlnIleGlySe 811
|||||
2335 AAGAAAAACCGGAAGACGAATCAACGCTTCCCAATTCCAACTCGGCAC 2384
811 rValPhe 813
2385 GACGTTT 2391
seq_name: gb_ba:AE002375

```

```

seq_documentation_block:
LOCUS AE002375 13538 bp DNA linear BCT 25-MAY-2000
DEFINITION Neisseria meningitidis serogroup B strain MC58 section 17 of 206 of
the complete genome.
ACCESSION AE002375 AE002098
VERSION AE002375.1 GI:7225394
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C.,
Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Peden, J.F.,
Dodson, R.J., Nelson, W.C., Gwinn, M.L., DeBoy, R., Peterson, J.D.,
Hickey, E.K., Haft, D.H., Salzberg, S.L., White, O., Fleischmann, R.D.,
Dougherty, B.A., Mason, T., Ciecko, A., Parksey, D.S., Blair, E.,
Cittone, H., Clark, E.B., Cotton, M.D., Utterback, T.R., Khouri, H.,
Qin, H., Vamathevan, J., Gill, J., Scarlato, V., Massignani, V.,
Pizza, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R.,
Rappuoli, R. and Venter, J.C.
Complete genome sequence of Neisseria meningitidis serogroup B
strain MC58
Science 287 (5459), 1809-1815 (2000)
20175755
10710307
REFERENCE
2 (bases 1 to 13538)
Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C.,
Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Peden, J.F.,
Dodson, R.J., Nelson, W.C., Gwinn, M.L., DeBoy, R., Peterson, J.D.,
Hickey, E.K., Haft, D.H., Salzberg, S.L., White, O., Fleischmann, R.D.,
Dougherty, B.A., Mason, T., Ciecko, A., Parksey, D.S., Blair, E.,
Cittone, H., Clark, E.B., Cotton, M.D., Utterback, T.R., Khouri, H.,
Qin, H., Vamathevan, J., Gill, J., Scarlato, V., Massignani, V.,
Pizza, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R.,
Rappuoli, R. and Venter, J.C.
Direct Submission
Submitted (17-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1. .13538
/organism="Neisseria meningitidis MC58"
/strain="MC58"
/db_xref="taxon:122586"
/note="serogroup: B"
complement(121. .1377)
/gene="NMB0176"
complement(121. .1377)
/gene="NMB0176"
/note="similar to GB:L02948 SP:P29011 PID:145703 GB:U00096
PID:1651591 percent identity: 76.98; identified by
sequence similarity; putative"

```

```

/codon_start=1
/transl_table=11
/product="D-amino acid dehydrogenase, small subunit"
/protein_id="AAF40633.1"
/db_xref="GI:7225395"
/translation="MKVLVLGAGVAGVSSAWYLAEAGHEVTVVIDRAEGVAMETSFANA
GQLSYGYTPPWAAPGIPTKALKWLFKSHPLPLFRPDGSLYOIEMQMLQNCQTAHYO
TKNRWVRISSEYREMFREFAQTGMNFRGKGTGIIQIFROTKEVAARQDQAVLERY
GVYRRLKPEECAPEPALARVTAKTAGGLHLPADATDCRLFTENLYLCOBKGVRF
HENQNTSRIDHNGLRKTKVETGREFEADAVVCAIAGCFSTVLAQDLNLPYVKGX
SLTLPVNSDGAIPVSTLDESAYKAITREDNRIRVCGMAELSGYEIKLEKRETLAL
VYNDLFPPEGGLDLSQASFWGSLRPMTDPSTPLIGTRFRDNLFLNLTGHTGLGWTMSLGS
KLTDIVSGKTEIRSDLSLSRYQA"
/gene="NMB0177"
complement(1456. .2847)
/gene="NMB0177"
complement(1456. .2847)
/note="similar to GB:AE000511 PID:2314080 percent
identity: 69.11; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="sodium/alanine symporter, putative"
/protein_id="AAF40634.1"
/db_xref="GI:7225396"
/translation="MQVFLDNPKAFETVSGWVWGPLMLMLLVGTGILLTVLLKGLQF
TMLGVALKOAFMPKPKKSGEGHGDISHFAALMTALSATIGTNTAGVATAVVTGSP
GAVFWNMTAIFCMATKYGCVLAVKYRVNNSKGSGPMYIEKGLGNKWKMAVA
FALFGTFASGIVSSVQSAQAVOTSGEIPAYTGITLVLTAVVGLGIGIARA
ASFIIPMAVYVYVGLGLSIIAINDALMPAVKLIFSDAFSAQAVAGAIGTVIRYGA
GVFVNEAGMSAPIAAAAKTDHPVRQALVMTGTFDITVVCSTGIVLMGLLGA
GGFVYKPEVSGAALTVTTFQKMLPGIGGWITVIGLFFAYSTILGWCYCEKCAVVF
GEKFAGLYRVGYVSSVMLCTVLSLDLVLWLASDTFNGLMALPNLIALLMLAKVIVNETR
DFKQKITNGELPH"
complement(3204. .3980)
/gene="NMB0178"
complement(3204. .3980)
/gene="NMB0178"
/note="similar to PID:1718489 percent identity: 100.00;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="acyl-(acyl-carrier-protein)--UDP-N-
acetylglucosamine O-acyltransferase"
/protein_id="AAF40635.1"
/db_xref="GI:7225397"
/translation="MTLIHPTAVIDPKAELDSGVKGVATVIGPNVQICANTEIGPHA
VINGHTSIGENNRIFQFASLGEIPQDKYRDEPTKLIIGNGTIRREFTFNLTGTGTGI
GETRIDDDNIMAYCHLADCVIGNHTIFANNASLAGHYTIGDYVVLGYTLVVFQCR
IGDYAMTAPAAGVHKDVPYFMAAGYRAEPAGINSEGMRRNGFTABQISAVKDVYKTL
YHRGIPFEAKADILRRATQALAVFRDFAQSARGIIR"
complement(4067. .4516)
/gene="NMB0179"
complement(4067. .4516)
/gene="NMB0179"
/note="similar to PID:1718488 percent identity: 100.00;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="(3R)-hydroxymyristoyl-(acyl carrier protein)
dehydratase"
/protein_id="AAF40636.1"
/db_xref="GI:7225398"
/translation="MDVQLPIEAKDIOKLIHPHYLPFLQDRITAFEPKMTLTAKNVT
INEPQGFHFDPDLVPTSGVRLITFANMAQACGTALISEGGRKENEFFFFAGIDEARFR
QVIPGQDLVFEVELTSPRRGIGKGNVAKVADGVQAVAEAIMCAKRVV"
complement(4550. .5596)
/gene="NMB0180"
complement(4550. .5596)
/gene="NMB0180"
/note="similar to PID:1718487 percent identity: 100.00;
identified by sequence similarity; putative"
/codon_start=1

```











```
458 ThrGlyLysHisValAsnAlaSerPheArgSerGluThrArgGluVa 474
|||||  ....  |||||||  ....  |||||||  ....  |||
179453 ACGGCAAGTCGGCCACATGCGCGCTCCAGAGCAAAACACCGCTTAA 179404
474 lTyrSerLeuGlyMetThrAsnProTyrPheThrValAsnGlyValSerG 491
|||||  ....  |||||||  ....  |||||||  ....  |||
179403 CGGCTCGCTGCTGTTACTGACCGCTACTTACGGCAGACGGGGTCAGCC 179354
491 lnsrLeuSerGlyTyrTyrArg.....LysThrLysTyrAspAsnLys 505
....  ....  |||  ....  ....  ....  ....  ....
179353 TGGGCTACCATGTTTACGAAAGCCTTCGACCCGCGCAAGCATCGACC 179304
506 AsnIleSerAsnTyrValLeuAspSerTyrGlyGlySerLysTyrG 522
|||||  ....  |||  ....  ....  ....  ....  ....
179303 AGCATCAACATATATAAACACCACCGCAGCGCAGGATCCGCATGAG 179254
522 yTyrProIleAspGluAsnGlnArgIleSerPheGlyLeuAsnAlaAspA 539
:  |||||  |||  ....  |||||  ....  |||||  ....  |||
179253 CGTGCCTGTTACCGAATACGACCGCGTGAATTCGGTTTGGTGCGAGAAC 179204
539 snThrLysLeuHisGlyGlyArgPheMetGlyIleSerAsnValLysGln 555
:  ....  ....  ....  ....  ....  ....  ....  ....
179203 ACCTGACCGCTCAAC.....ACCTACCAACAAAGCGCCCAA 179169
556 LeuMetAlaAspGlyGlyLysIleGlnValAspAsnAsnGlyIleProAs 572
|||||  |||  ....  ....  ....  ....  ....  ....
179168 CACTATGCCGACTTTATCAAGAATAACGCCAAACCGACGGC....ACAGA 179122
572 pPheLysHisAspTyrThrTyrAsnAlaIleLeuGlyTyrAsnTyrS 589
|  ....  ....  ....  ....  ....  ....  ....  ....
179121 CGCGCACTCAAAAGCTGGCTGTACAAAGTACCGTCGGTCGGGCGCA 179072
589 eSerLeuAspArgProValPheProThrGlnGlyMetSerHisServa 605
:  ....  |||  ....  ....  ....  ....  ....  ....
179071 ACAAAACCGACACGCGGTATGGCCACCGCGCGCTACCTAGCGGCGG 179022
606 AspLeuThrValGlyPhe...ThrHisGlyLysValVa 620
:  ....  ....  ....  ....  ....  ....  ....  ....
179021 AACGGCGAAATCGCCCTGCTGGCAGCAAACTCACTACTCCGCCAC 178972
620 lTyrGlnGlyAsnIleTyrArgProPheIleLysLysSerValLeuArg. 636
:  ....  ....  ....  ....  ....  ....  ....  ....
178971 CCACAACCAACCTGTTCTCCCTGTAGCAAAACCTTCACGCTGATGC 178922
637 .....GlyTyrAlaLysLeuGlyTyrGly.....AsnAsn 646
178921 TCGGGCGCGAAGTCGCATTCGGGCG....GGCTACGGCAGAACCAAGAA 178875
647 LeuProPheTyrGluAsnPheTyrAlaGlyGlyTyrGlySerValArgG 663
:  |||||  ....  |||||  ....  |||||  ....  |||||  ....
178874 ATCCCTCTCTTGAAAACCTTACGCGCGCGGCTGGGTGCGTGGCGG 178825
663 yTyrAspGlnSerSerLeuGlyProArgSerGlnAlaTyrLeuThrAla 680
|||||  ....  |||||  ....  ....  ....  ....  ....
178824 ATACGAAGCGGCAGCTCGGTCGGA..... 178797
680 rgArgGlyGlnGlnThrLeuGlyGluValVal.....GlyGlyAsn 694
|||||  ....  |||||  ....  ....  ....  ....  ....
178796 .....GTCTATGACGAATACGGCGAAATAATCAGCTACGCGCGCAAC 178755
695 AlaLeuAlaThrPheGlySerGluLeuIleLeuProLeuProPheLysG 711
|||||  ....  |||||  ....  ....  ....  ....  ....
178754 AAAAAAGCCCAAGCTCTCCGCGAGCTGCTTCCCGATCCCGCGCGCGAA 178705
711 yAspTrpIleAspGlnValArgProValIlePheIleGluGlyGlyGln 728
|||  ....  ....  ....  ....  ....  ....  ....
178704 AGAC...GCGCGCACCGCTCGCGCTGAGCTGTTTCCGCGACGAGGCGAG 178658
728 aPheAspThrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPhe 744
(|:|:|)  |||  ....  ....  ....  ....  ....
178657 TGTGGACGCGCAAAACCTACGACGACACAGCAGATTCGCGCAGCC..... 178614
```

```
745 LysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgPr 761
:  ....  ....  ....  ....  ....  ....  ....  ....
178613 .....GGCGGAGGGTTCAAACATTTACGGCGCGCAATACCCA 178573
761 oLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaT 778
|||  ....  |||||  ....  |||||  ....  |||||  ....
178572 TAAATCCACTTTACCAACGAATTCGCTATTCGCGCGCGCGCGGTTA 178523
778 hrTrpTyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeu 794
|||||  ....  |||||  ....  ....  ....  ....  ....
178522 CTFGGCTCTCGCCTTTAGGCCCGATGAATTCAGTACGCTACCCGCTG 178473
795 AsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGlnIleGlySe 811
:  |||||  ....  ....  ....  ....  ....  ....
178472 AAGAAAAACCGGAGACGAAATCCACGCTTCCAATTCCACTCGGCAC 178423
811 rValPhe 813
:  ....  ....
178422 GACGTTC 178416
seq_name: gb_pat:AX155447
seq_documentation_block:
LOCUS AX155447 2394 bp DNA linear PAT 22-JUN-2001
DEFINITION Sequence 10 from Patent WO0138350.
ACCESSION AX155447
VERSION AX155447.1 GI:14536772
KEYWORDS
SOURCE Neisseria gonorrhoeae.
ORGANISM Neisseria gonorrhoeae
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
Neisseria.
REFERENCE 1 (bases 1 to 2394)
AUTHORS Giuliani M.M., Pizsa M., Rappuoli R. and Holst J.
TITLE 85kda neisserial antigen
JOURNAL Patent: WO 0138350-A 10 31-MAY-2001;
Chiron Spa (IT); Statens Institut for Folkehelse (NO)
FEATURES
source 1. 2394
/organism="Neisseria gonorrhoeae"
/db_xref="taxon:485"
BASE COUNT 655 a 758 c 586 g 395 t
ORIGIN
alignment_scores:
Quality: 1111.00 Length: 819
Ratio: 2.035 Gaps: 14
Percent Similarity: 66.667 Percent Identity: 32.357
alignment_block:
US-09-701-711-2 x AX155447 ..
Align seg 1/1 to: AX155447 from: 1 to: 2394
10 GlnValSerAlaMetThrMetAlaValMetMetValMetSerThrHisAl 26
:  ....  ....  ....  ....  ....  ....  ....
4 AAACCTGAACAGATTCGTCCTCCGACATGATGCTTGGGCATATCGCCTTT 53
26 aGlnAlaAlaAspPheMetAlaAsnAspIleThrIleThrGlyLeuGlnA 43
|||||  ....  |||||  ....  ....  ....  ....
54 GGCACCTGCGGACTTCACCATCCAGACATCCGCGTCGAAGGCTTGCGAG 103
43 rgValThrIleGluSerLeuGlnSerValLeuProPheArgLeuGlyGln 59
|||||  ....  ....  ....  ....  ....  ....
104 GTACCGAGCGGAGTACCTACCTATTCACACTCGCTCCGCTCAAGTCGCG 153
60 ValValSerGluAsnGlnLeuAlaAspGlyValLysAlaLeuTyrAlaTh 76
:  ....  ....  ....  ....  ....  ....
154 ACCTACACGACACACACGCGAGTGCATCATCAAAGCCTGTACGCCAC 203
76 rGlyAsnPheSerAspValGlnValTyrHisGlnGlyArgIleIleT 93
|||||  |||||  ....  ....  ....  ....  ....
```



```

680 rgargSLyngInthrThrleuGLyIuValVal.....GLYGLYAsn 694
2011 .....GTGATATGCAATACGGCGAAAAATCATCAGCTACGGCGGCAC 2052
695 AAlaLeuAlaThrPheGLySerGLuLeuLeuProLeuProPheLysGL 711
2053 AAAAAAGCCACAGCTCCGCCGAGCTGCTTCCGATGCCCGCGCGCA 2102
711 yAspTrpIleAspGLInValArgProValIlePheIleGLyGLyGLn 728
2103 AAGC...GGCGCAGCCGCTCCGCTGAGCTGTTGGCGAGCGAGGAGCG 2149
728 aLpheAspThrThrGLyMetAspLysGLInThrIleAspLeuThrGLn 744
2150 TGTGGACGGCAAAACCTACGACGACACAGCAGTTCGGCGACC..... 2193
745 LysAspProGLInAlaThrAlaGLuGLnAsnAlaLysAlaAsnArgPr 761
2194 .....GGCGGCGAGGTTCAAAACATTTCAGCGCGCGCAATGCCA 2234
761 oLeuLeuThrGLnAspLysGLInLeuArgTyrSerAlaGLyValAla 778
2235 TAAATCCACCTTTACCAACGAATTCGCTATTCGCCGCGCGCGCTTA 2284
778 hrTrpTyrThrProIleGLyProLeuSerIleSerTyrAlaLysPro 794
2285 CCTGGCTCTCGCTTTCAGCCGATCAAAATTCAGCTACGCTACCGCTG 2334
795 AsnLysLysGLInAsnAspGLInThrAspThrValGLnPheIleGL 811
2335 AAGAAAAACCGGAAGACGAATCCAAACGCTTCACATTCCAACTCGGC 2384
811 rValPhe 813
2385 GAGCTTC 2391

seq_name: gb_pat:AX202497

seq_documentation_block:
LOCUS AX202497 2394 bp DNA linear PAT 30-AUG-2001
DEFINITION Sequence 10 from Patent WO0152885.
ACCESSION AX202497
VERSION AX202497.1 GI:15392217
KEYWORDS
SOURCE
ORGANISM
Neisseria meningitidis.
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
Neisseria.
REFERENCE
1 (bases 1 to 2394)
AUTHORS
Plaza,M.C., Rappuoli,R.C. and Giuliani,M.C.
TITLE
Outer membrane vesicle (omv) vaccine comprising N. meningitidis
serogroup b outer membrane proteins
Patent: WO 0152885-A 10 26-JUL-2001;
JOURNAL
Chiron Spa (IT)
FEATURES
source
1..2394
Location/Qualifiers
BASE COUNT 655 a 758 c 586 g 395 t
ORIGIN
/description="Neisseria meningitidis"
/db_xref="taxon:487"

alignment_scores:
Quality: 1111.00 Length: 819
Ratio: 2.035 Gaps: 14
Percent Similarity: 66.667 Percent Identity: 32.357

alignment_block:
US-09-701-711-2 x AX202497 ..
Align seg 1/1 to: AX202497 from: 1 to: 2394
10 GlnValSerAlaMetThrMetAlaValMetMetValMetSerThrHisAl 26

```

```

.....: .....: |||.....:
4 AACTGAAACAGATGTGCTCCGACATGATGTCTTGGGCAATATCGCCTT 53
26 agInAlaAlaAspPheMetAlaAsnAspIleThrIleThrGLyLeuGln 43
54 GGCACCTTCGCGACTTCACCATCCAGACATCCGCTGCAAGCGCTTCAGC 103
43 rGLValThrIleGLuSerLeuGLnSerValLeuProPheArgLeuGLn 59
104 GTCACGAGCCGAGTACCGATTCATTCACACTACTGCCCGTCAAAAGTGGGAG 153
60 ValValSerGLuAsnGLnLeuAlaAspGLyValLysAlaLeuTyrAla 76
154 ACTTACACAGCACACACAGCGCACTGCCATCATCAAAACCTGTACGCGAC 203
76 rGLYAsnPheSerAspValGLnValTyrHisGLnGLyArgIleLeu 93
204 CGGTTCTTTCGACGAGTACGCTCGAACTGGGAGCGGACACTCTCTGC 253
93 yrGLnValThrGLuArgProLeuIleAlaGLuIleAsnPheGLuGLyAsn 109
254 TGACCGTTATGACAGCCGCCACCATCGGCTCGTCAACATCACCGCGCA 303
110 ArgLeuIleProLysGLyLeuGLnGLyLeuLysAsnAlaGLy 126
304 AAATGCTGCAAAACGACGCCATTAAAGAAAACCTGCATCGTTCGGCT 353
126 uAlaValGLyGLnProLeuLysGLnAlaThrValGLnMetIleGLuThr 143
354 GCGCGAGTCCGCAATCTTATTCAGCGGACACATCAACGACGAGTCCGC 403
143 IuLeuThrAsnGLnTyrIleSerGLnGLyTyrAsnThrGLuIleThr 159
404 GCCTGAAGAAAGATACCTCGGCGCGCAACTATATATCAATATCAG 453
160 ValLysGLnThrMetLeuAspGLyAsnArgValLysLeuAspMetThr 176
454 CCCAAAGTACCAAACTCGCCCGCAACCGGCTGACATCGACATCGACAT 503
176 eAlaGLyGLyLysProAlaArgValAlaAspIleAsnIleIleGLyAsn 193
504 TGACGAGGGCAAAATCCGCAAAATTCACCGCATTCGAATTTGAAGGCAAC 553
193 IuHisPheSerAspAlaAspLeuIleAspValLeuAlaIleLysAspAsn 209
554 AACTCTATTCGACCGCAAACTGATGCGGAGATGTCGTCGACCGAAGGC 603
210 LysIle...AsnProLeuSerLysAlaAspArgTyrThrGLnGLuLys 225
604 GGCATTGGACATGGCTGACACACAGCAACCAATTCACAGCAGCAAAAT 653
225 uValThrSerLeuGLuAsnLeuArgAlaLysTyrLeuAsnAlaGLyPhe 242
654 TGCCCAAGACATGAAAGTAACCGACTTTCACGAAACACAGCGCTACT 703
242 aLArgPheGLuIleLysAspAlaLysLeuAsnIleAsnGLuAspLysAsn 258
704 TCGATTTCGCAATCTCGATACGACATCCAAACCAAGCAAGCAAAAC 753
259 ArgIlePheValGLuIleSerLeuHisGLuGLyGLuIleTyrArgPhe 275
754 AAGCAGACCATCAAAATCACCGCTCCACGAAAGCGGAGCTTCGCTGGG 803
804 CAAGTCTCATGGAAGCGACCAACCAAGCAAGTCCCAAGCGCAACTCG 853
291 IuAlaLeuLeuLysPheLysAlaGLuGLuGLyPheSerGLnAlaMetLeu 307
854 AAAAAGCTGCTGACCATGAAGCCCGCAAAATGTATGACAGACCGACGAT 903
308 GLuGLnThrThrAsnAsnIleSerThrLysPheGLyAspAspLysTyr 324
.....: .....: |||.....:

```

904 ACCGCCGTTTGGGTGAGATTTCAGAACCGCATGGCGGCGAGGCTACGC 953  
 324 TTYTALAGlnIleArgProValThrArgIleAsnAspIleSerArgThrV 341  
 954 ATACAGCGAAATCAGCTACAGCCGCTGCCAAACCGCAACCAACCAACCG 1003  
 341 aAspValGlnTyrThrIleAspProValHisProValTyrValArgArg 357  
 1004 TCGATTTCGCTCGCATCGAACCGGCGCGGAAATCTACGTCACAGAA 1053  
 358 IleAsnPhetThrGlyAsnPhetThrGlnAspGluValIleuArgArgI 374  
 1054 ATCCACATCCGGGACACACAAACCGCGAGAGTGTGCGCGCGCA 1103  
 374 uMetArgGlnLeuGlnGluValAlaLeuAsnGlnIleGlnLeuS 391  
 1104 ATTGGCGCAATGGAATCCGGCGCTTACGACACCTCCAAAGTCGCAACGT 1153  
 391 eArgAlaArgLeuMetArgThrGlyPhePheIleHisValThrValAsp 407  
 1154 CCAAAGACGGCGGTGAGCTTTGGGCTTTCGACACGTCAGTTGTGAT 1203  
 408 ThrArgProValProAsnSerProAspGlnValAspValAsnPhetVal 424  
 1204 GCGTCGCGCTGGCGGACACCGACAAAGTCGATTGGAACATGACGCT 1253  
 424 IGluglnIleProSerGlySerSerThrIleAlaIleArgIleSerGln 441  
 1254 GACCGACGCTCCACCGGCTCGCTCGACTTGAGCGCGGCTGGTACAGC 1303  
 441 eArgIleValThrPheGlnPheAspValSerGlnAsnAsnPhetGly 457  
 1304 ATACCGGCTGGTGTGTCGCGAGGCTTCCAGACAACTGTGCTGCT 1353  
 458 ThrGlyHisValAsnAlaSerPheSerArgSerGluThrArgGluVal 474  
 1354 ACGGCAAGTGGCGCGCTCGCGGCTCAGCAACCAACCAACCTCA 1403  
 474 ITyrSerLeuGlyMetThrAsnProTyrPheThrValAsnGlyValSer 491  
 1404 CGGCTCGCTGTGTACCGACCGCTACTTCAGCGCAGCGGGTGAGCC 1453  
 491 IAsnLeuSerGlyTyrTyrArg....LysThrLysTyrAspAsnLys 505  
 1454 TGGGCTACGATGTTTACGAAAGCTTCGACCGCGCAACCAATCGAC 1503  
 506 AsnIleSerAsnTyrValLeuAspSerTyrGlySerLeuSerTyrG 522  
 1504 AGCATCAACATATATAAACCAACCGCAGCGCAGCATCCGATGAG 1553  
 522 YThrProIleAspGlnAsnGlnArgIleSerPheGlyLeuAsnAlaAsp 539  
 1554 CGTGCTGTTCACGAATACGACCGGCTGTAATTCGTTGGTGCGAGAAC 1603  
 539 SerThrLysLeuHisGlyLysArgPheMetGlyIleSerAsnValLysGln 555  
 1604 ACCTGACCGCTCAAC.....ACCTACACAAACCGCCCAAA 1638  
 556 LeuMetAlaAspGlyLysIleGlnValAspAsnAsnGlyIleProAs 572  
 1639 CACTATGCGGACTTTATCAAGAAATACGGCAAAACCGAGGCG...ACAGA 1685  
 572 PPhelYHisAspTyrThrThrTyrAsnAlaIleLeuGlyTyrPAsnTyr 589  
 1666 CGCGAGCTTCAAGAGCTGCTCTACAAAGTACCTCGCTGGGCGGCGA 1735  
 589 eArgLeuAspArgProValPheProThrGlnGlyMetSerHisSerVal 605  
 1736 ACAAAACGACAGCGCTTATGCGCGCGGCTACTACTGACGGGCGTG 1785  
 606 AspLeuThrValGlyPhe...GlyAspLys...ThrHisGlnLysValVa 620  
 1786 AACGCCGAATCGCCCTGCGCGGCAAAACCTGCAATACTACTCGCCAC 1835

620 ITyrGlnGlyAsnIleTyrArgProPheIleLysSerValLeuArg. 636  
 1836 CCACAAACCAACCTGCTCTTCCCTTAAGCAAAACCTTACAGCTGARGC 1885  
 637 .....GlyTyrAlaLysLeuGlyTyrGly.....AsnAsn 646  
 1886 TCGCGCGCGCAACCTCGCATTTGGGGGC...GGCTACGGGAGAAACCAAA 1932  
 647 LeuProPheTyrGluAsnPhetThrAlaGlyLysTyrGlySerValArgI 663  
 1933 ATCCCTTCTTGAATACTTCTACGGCGCGGCTGGGCTGGTGGCGCGG 1982  
 663 YThrAspGlnSerLeuGlyProArgSerGlnAlaTyrLeuThrAla 680  
 1983 ATACGAAGCGGACGCTGCTGTCGAAA..... 2010  
 680 rGArgGlyGlnIleThrThrLeuGlyGluValVal.....GlyGlyAsn 694  
 2011 .....GTGTATGACGAATATGCGGAAATATGATGCTACGCGCGCAAC 2052  
 695 AlaLeuAlaThrPheGlySerGlnLeuIleuProLeuProPheLysG 711  
 2053 AAAAAAGCAAGCTCTCGCGGAGCTGCTTCCGATGCGCGGCGCGCA 2102  
 711 YAspThrIleAspGlnValArgProValIlePheIleGlnGlyGlnV 728  
 2103 AGAC...GGCGCACCGCTCGCGCTGAGCTGTTCGCGAGGAGCAGCG 2149  
 728 aLPhetAspThrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPhe 744  
 2150 TGTGGAGCGCAAAACCTTACGAGCAACAGCATTCGCGCAC.... 2193  
 745 LysAspProGlnAlaThrAlaGlnIleAsnAlaTyrAlaAlaAsnArgPr 761  
 2194 .....GGGCGCAGGCTTCAAAACATTTACGGCGCGCGGCAATACCA 2234  
 761 OleuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaT 778  
 2235 TAAATCCACCTTACCAACGAATTCGCTATTCGCCGCGCGGCGGTGA 2284  
 778 hTriPyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeu 794  
 2285 CTGGCTCTCGCTTATGAGCCGATGAATTCAGCTACGCTCAACCGCTG 2334  
 795 AsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGlnIleGlySe 811  
 2335 AAGAAAAACCGGAGAGCAATCAACGCTTCCAAATCCAACTCGGCGAC 2384  
 811 rValPhe 813  
 2385 GACGTTTC 2391  
 seq\_name: gb\_ba:NM12491  
 seq\_documentation\_block: 340806 bp DNA linear BCT 04-DEC-2000  
 LOCUS NM12491  
 DEFINITION Neisseria meningitidis serogroup A strain Z2491 complete genome;  
 segment 1/7.  
 ACCESSION AL162752 AL157959  
 VERSION AL162752.2 GI:7378778  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Neisseria meningitidis Z2491.  
 Neisseria meningitidis Z2491  
 Bacteria; proteobacteria; beta subdivision; Neisseriaceae;  
 Neisseria.  
 REFERENCE  
 1 (bases 1 to 340806) James, R.D., Bentley, S.D., Churcher, C.,  
 Parkhill, J., Achtman, M., Klee, S.R., Morelli, G., Basham, D., Brown, D., Chillingworth, T.,  
 Davies, R.M., Davis, P., Devlin, K., Feltham, T., Hamlin, N.,  
 Holtroyd, S., Jags, K., Leather, S., Moul, S., Mungall, K.,  
 Quail, M.A., Rajandream, M.A., Rutherford, K.M., Simmonds, M.,  
 Skelton, J., Whitehead, S., Spratt, B.G. and Barrall, B.G.

Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491  
 JOURNAL Nature 404 (6777), 502-506 (2000)  
 MEDLINE 20222556  
 REFERENCE 2 (bases 1 to 340806)  
 AUTHORS Parkhill, J.  
 TITLE Direct Submission  
 JOURNAL Submitted (30-MAR-2000) Submitted on behalf of the Neisseria sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk

Notes:  
 Details of *N. meningitidis* sequencing at the Sanger Centre are available on the world wide web.  
 (URL, [http://www.sanger.ac.uk/Projects/N\\_meningitidis/](http://www.sanger.ac.uk/Projects/N_meningitidis/)).

FEATURES

source	1..340806 /organism="Neisseria meningitidis Z2491" /strain="Z2491" /db_xref="taxon:122587" /note="serogroup: A" complement(38..358) /gene="NMA0001" complement(38..358) /gene="NMA0001" /codon_start=1 /product="hypothetical protein NMA0001" /protein_id="CAB83321.1" /db_xref="GI:7378779" /db_xref="SPTREMBL:Q9JX93" /translation="MSNTQIRSADEFTTSRAWGALDIANNMGTTVRLHWTDDQPKYHV NDGEVFAVMGDEVMDHYRENGEEHIVRLKSGDIFVAGICTEHAHPGETRILIVIEK EGSV" complement(414..2438) /gene="nuol" complement(414..2438) /gene="nuol"
gene	674 aa; similar to many e.g. NUOL_RHOCA NADH dehydrogenase I chain L (EC 1.6.5.3) (700 aa). Contains Pfam match to entry PF00361 oxidored_q1, NADH-Ubiquinone/plastoquinone (complex I), various chains, and to entry PF00662 oxidored_q1.N, NADH-Ubiquinone oxidoreductase (complex I), chain 5 N-terminus"
CDS	/codon_start=1 /transl_table=11 /product="hypothetical protein NMA0001" /protein_id="CAB83321.1" /db_xref="GI:7378779" /db_xref="SPTREMBL:Q9JX93" /translation="MSNTQIRSADEFTTSRAWGALDIANNMGTTVRLHWTDDQPKYHV NDGEVFAVMGDEVMDHYRENGEEHIVRLKSGDIFVAGICTEHAHPGETRILIVIEK EGSV" complement(414..2438) /gene="nuol" complement(414..2438) /gene="nuol"
misc_feature	1035..1044 /note="Core DNA uptake sequence: gccgtctgaa" /label=DUS
misc_feature	complement(1116..2030) /gene="nuol" /note="Pfam match to entry PF00361 oxidored_q1, NADH-Ubiquinone/plastoquinone (complex I), various chains, score 351.80, E-value 7.3e-102"
misc_feature	complement(2061..2243) /gene="nuol"

gene	/gene="nuol" /note="Pfam match to entry PF00662 oxidored_q1.N, NADH-Ubiquinone oxidoreductase (complex I), chain 5 N-terminus, score 77.20, E-value 3.5e-19" complement(2487..2828) /gene="NMA0003" complement(2487..2828) /gene="NMA0003" /note="NMA0003, unknown, len: 113 aa" /codon_start=1 /transl_table=11 /product="hypothetical protein NMA0003" /protein_id="CAB83323.1" /db_xref="GI:7378781" /db_xref="SPTREMBL:Q9JX91" /translation="MMDKNQLEQEFHFKAMLIYOEALNLPQYKATRFLOIVNEFGGK EADKLLSTGSEKKTQTGTETLILSGGVHALKSMYELVLQKPCWDLFTBQLAVARK RLKRVGFVSPK" complement(2833..2837) complement(2952..3527) /gene="NMA0004" complement(2952..3527) /gene="NMA0004" /note="NMA0004, unknown, len: 191 aa; similar to hypothetical proteins e.g. Y977_HAEIN HI0977 (191 aa), fasta scores: E(): 0; 80.5% identity in 190 aa overlap, and to proteins involved in cell cycle e.g. TR-O54679 (EMBL:AF036487) Lactococcus lactis Plasmid pNZ4000 putative mobilization protein (200 aa), fasta scores: E(): 0, 55.1% identity in 167 aa overlap, and FIC_ECOLI cell filamentation protein FIC (200 aa), fasta scores: E(): 0.093, 29.4% identity in 126 aa overlap" /codon_start=1 /transl_table=11 /product="hypothetical protein NMA0004" /protein_id="CAB83324.1" /db_xref="GI:7378782" /db_xref="SPTREMBL:Q9JOR9" /translation="MPSENPIGTMKSIDQSLNARRLFESGDIDRIVGCTAGLQO IHRYLGGLYDFAGQIREDNISKGGFRANAMYLKEALVKIEQNPETFEELIAKYVE MNAPLEPGNGRSTRWLVLNKKVKVYVQNVSKTLYLQAWERSPVNDLELRLF LKNDLTDVDNREIIFKGIQSYIEGYEGK" complement(3516..3525) /gene="NMA0004" /note="Core DNA uptake sequence: gccgtctgaa" /label=DUS complement(3561..3866) /gene="nuok" complement(3561..3866) /gene="nuok" /EC_number="1.6.5.3" /note="NMA0005, nuok, NADH dehydrogenase I chain K, len: 101 aa; similar to many e.g. NUOK_RHOCA NADH dehydrogenase I chain K (EC 1.6.5.3) (102 aa), fasta scores: E(): 1.1e-24; 65.0% identity in 100 aa overlap. Contains Pfam match to entry PF00420 oxidored_q2, NADH-ubiquinone/plastoquinone oxidoreductase chain 4L" /codon_start=1 /transl_table=11 /product="NADH dehydrogenase I chain K" /protein_id="CAB83325.1" /db_xref="GI:7378783" /db_xref="SPTREMBL:Q9JOJ9" /translation="MITLTHVLVIGALLFGISAMGIFMRNKNVLVLMSTELMLAVN FNFIAPSOHLGDTAGQIFVFVLTVAANESAIGLAINVLYVRNQTINVADLDELKG" complement(3564..3863) /gene="nuok" /note="Pfam match to entry PF00420 oxidored_q2, NADH-ubiquinone/plastoquinone oxidoreductase chain 4L, score 141.50, E-value 1.5e-38" complement(3863..4534) /gene="nuoj" complement(3863..4534)
misc_feature	
gene	
CDS	
misc_feature	
gene	
CDS	

```

/EC_number="1.6.5.3"
/Note="NMA0006, nuoJ, NADH dehydrogenase I chain J, len:
223 aa; similar to many e.g. NUOJ, RHOCA NADH dehydrogenase
I chain J (EC 1.6.5.3) (202 aa), fatsa scores: E():
9, 7e-21; 41.3% identity in 201 aa overlap. Contains Pfam
match to entry PF00499 oxidored_q3,
NADH-ubiquinone/plastoquinone oxidoreductase chain 6"
/codon_start=1
/transl_table=11
/product="NADH dehydrogenase I chain J"
/protein_id="CAB83326.1"
/db_xref="GI:7378784"
/db_xref="SPTREMBL:Q9JX90"
/translation="MTFGLIFFYFAVILYXGAIKTVAKNPVHAALHLVTFVCSAM
LWMLQAEFLGVTLVVVVGVAVVLFVWMLNIDIEEMRAGFWRHAPVAGVVGTL
AVALLILVNPKTDLAAFLMKMDIPADYNNIRLGSRTYDYLLPFLAAVLLLGWV
RAIALVHRKTVNPKRMDPADQVKVRADQGRMLVKMEAVKPTQTESAESEVSDLLKPK
EEGKA"
RBS
complement(3874..3879)
/EC_number="1.6.5.3"
/Note="NMA0006, nuoJ, NADH dehydrogenase I chain J"
/protein_id="CAB83326.1"
/db_xref="GI:7378784"
/db_xref="SPTREMBL:Q9JX90"
/translation="MTFGLIFFYFAVILYXGAIKTVAKNPVHAALHLVTFVCSAM
LWMLQAEFLGVTLVVVVGVAVVLFVWMLNIDIEEMRAGFWRHAPVAGVVGTL
AVALLILVNPKTDLAAFLMKMDIPADYNNIRLGSRTYDYLLPFLAAVLLLGWV
RAIALVHRKTVNPKRMDPADQVKVRADQGRMLVKMEAVKPTQTESAESEVSDLLKPK
EEGKA"
misc_feature
complement(4034..4525)
/EC_number="1.6.5.3"
/Note="NMA0006, nuoJ, NADH dehydrogenase I chain J"
/protein_id="CAB83326.1"
/db_xref="GI:7378784"
/db_xref="SPTREMBL:Q9JX90"
/translation="MTFGLIFFYFAVILYXGAIKTVAKNPVHAALHLVTFVCSAM
LWMLQAEFLGVTLVVVVGVAVVLFVWMLNIDIEEMRAGFWRHAPVAGVVGTL
AVALLILVNPKTDLAAFLMKMDIPADYNNIRLGSRTYDYLLPFLAAVLLLGWV
RAIALVHRKTVNPKRMDPADQVKVRADQGRMLVKMEAVKPTQTESAESEVSDLLKPK
EEGKA"
gene
complement(4570..5277)
/EC_number="1.6.5.3"
/Note="NMA0006, nuoJ, NADH dehydrogenase I chain J"
/protein_id="CAB83326.1"
/db_xref="GI:7378784"
/db_xref="SPTREMBL:Q9JX90"
/translation="MTFGLIFFYFAVILYXGAIKTVAKNPVHAALHLVTFVCSAM
LWMLQAEFLGVTLVVVVGVAVVLFVWMLNIDIEEMRAGFWRHAPVAGVVGTL
AVALLILVNPKTDLAAFLMKMDIPADYNNIRLGSRTYDYLLPFLAAVLLLGWV
RAIALVHRKTVNPKRMDPADQVKVRADQGRMLVKMEAVKPTQTESAESEVSDLLKPK
EEGKA"
CDS
complement(4570..5277)
/EC_number="1.6.5.3"
/Note="NMA0006, nuoJ, NADH dehydrogenase I chain J"
/protein_id="CAB83326.1"
/db_xref="GI:7378784"
/db_xref="SPTREMBL:Q9JX90"
/translation="MTFGLIFFYFAVILYXGAIKTVAKNPVHAALHLVTFVCSAM
LWMLQAEFLGVTLVVVVGVAVVLFVWMLNIDIEEMRAGFWRHAPVAGVVGTL
AVALLILVNPKTDLAAFLMKMDIPADYNNIRLGSRTYDYLLPFLAAVLLLGWV
RAIALVHRKTVNPKRMDPADQVKVRADQGRMLVKMEAVKPTQTESAESEVSDLLKPK
EEGKA"

```

## alignment\_scores:

```

Quality: 1111.00 Length: 819
Ratio: 2.035 Gaps: 14
Percent Similarity: 66.667 Percent Identity: 32.357

alignment_block:
US-09-701-7111-2 x NMA122491 ..
Align seg 1/1 to: NMA122491 from: 1 to: 340806

10 GlnValSerAlaMetThrMetAlaValMetMetValMetSerThrHisAl 26
83985 AAACGTGAACAGATGGCTCCGACATGATGGCTTGGGCGATATCGCCTTT 84034
26 aGlnAlaAspPheMetAlaAsnAspIleThrIleThrGlyLeuGlnA 43
84035 GGCACCTTGGCGACTTCACCATCCAGACATCCGCGTCGAAGGCTGCGAGC 84084
43 rgValThrIleGluSerLeuGlnSerValLeuProPheArgLeuGlyGln 59
84085 GTACCGAGCGAGTACCGGTATTCACCTACCTGCGCGTCAAAAGTCGGCGAC 84134
60 ValValSerGluAsnGlnLeuAlaAspGlyValValValValValAlaTh 76
84135 ACCTACACGACACACAGCGAGTGCATCAAAAGCCTGTACGCGCAC 84184
76 rGlyAsnPheSerAspValGlnValTyThrHisGlnGluGlyArgIleIle 93
84185 CGGTTTCTTTCAGCAGCTACGCGTGAACACTCGCGAGCGGCGAGCTCTCGC 84234
93 yrGlnValThrGluArgProLeuIleAlaGluIleAsnPheGluGlyAsn 109
84235 TGACCGGTATTCGAACGCGCCACCATCGCTCGCTCAACATCACCGGCGCA 84284
110 ArgLeuIleProLysGluGlyLeuGlnGluGlyLeuLysAsnAlaGlyLe 126

```

```

84285 AAATGTCGCAACAGCGCATTAAAGAAAACCTCGAATCGTTTCGGGCT 84334
126 uAlaValGlyGlnProLeuLysGlnAlaThrValGlnMetIleGluThrG 143
84335 GCGCAGCTCGCAATCTTAAATCAGGCGACACTCAATCAGCAGCTCGCG 84384
143 luLeuThrAsnGlnTyIleSerGlnGlyTyTyAsnThrGluIleThr 159
84385 GCCTGAAGAAGAAATACCTCGGCGCGCAAACTCAATATCAATCACG 84434
160 ValLysGlnThrMetLeuAspGlyAsnArgValLysLeuAspMetThrPh 176
84435 CCCAAATACCAAACTCGCCGCAACCGCTCGACATCGACATCACGAT 84484
176 eAlaValGlyGlyProAlaArgValAlaAspIleAsnIleIleGlyAsnG 193
84485 TGACGAGGCGAAATCCGCCAAATATCCGACATCGAATTTGAAGCAACC 84534
193 lnHisPheSerAspAlaAspLeuIleAspValLeuAlaIleLysAspAsn 209
84535 AAGTCTATTCGACCGCAAACTGATGGCGAGATGCTGCTACCGCAAGGC 84584
210 LysIle...AsnProLeuSerLysAlaAspArgTyThrGlnGluLysLe 225
84585 GGCATTTGGACATGGCTGACGACGAAACCAATTCACGAGCAGAAAT 84634
225 uValThrSerLeuGluAsnLeuArgAlaLysTyThrLeuAsnAlaGlyPheV 242
84635 TGCCCAAGACATGGAAAAGTAACCGACTTCTACCAAGACACACGCTACT 84684
242 alArgPheGluIleLysAspAlaLysLeuAsnIleAsnGluAspLysAsn 258
84685 TCGATTTCCGCTCTCCTGATACCGACATCCAAACCAAGCAAGCAAAAC 84734
259 ArgIlePheValGluIleSerLeuHisGluGlyGluGlnTyArgPheG 275
84735 AAGCAGACCATCAAAATCACCGTCCAGAGGCGGACGTTTCCGTTGGG 84784
275 yGlnThrGlnPheLeuGlyAsnLeuThr...TyThrGlnAlaGluLeuG 291
84785 CAAGTCTCCATCGAAGCGCACCAACGAAGTCCCAAGCAACGCACTGG 84834
291 luAlaLeuLysPheLysAlaGluGlyPheSerGlnAlaMetLeu 307
84835 AAAAATCTGTGACCATGAACCGCGCAATGGTACGAACGCCACGACATG 84884
308 GluGlnThrThrAsnAsnIleSerThrLysPheGlyAspAspGlyTyTy 324
84885 ACCGCCGTTTTGGGTGAGATTACAGACCGCATGGGCTCGGCGAGCTAGC 84934
324 rTyAlaGlnIleArgProValThrArgIleAsnAspGluSerArgThrV 341
84935 ATACAGCAAAATCAGCGCTACAGCGCTGCCAAAGCGCCGCAACCAACCG 84984
341 alAspValGluTyTyIleAspProValHisProValTyValArgArg 357
84985 TCGATTCGCTCTGACATGAACCGCGCGGCAAAATCTACGTCACAGAA 85034
358 lIleAsnPheThrGlyAsnPheLysThrGlnAspGluValLeuArgG 374
85035 ATCCACATACCGGCAACACAAACCGCGGACGAGTCTGTCGCGCGCGA 85084
374 uMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnLeu 391
85085 ATTGCGCAAAATGGAATCCGCGCTTACGACACCTCCAAAGCTGCAACGCT 85134
391 erArgAlaArgLeuMetArgThrGlyPhePheLysHisValThrValAsp 407
85135 CCAAGAGCGCGCTGAGCTTTTGGCTTCTGACACAGTACAGTACGTTTGCAT 85184
408 ThrArgProValProAsnSerProAspGlnValAspValAspValValVa 424

```









```
10  GlnValSerAlaMetThrMetAlaValMetMetValMetSerThrHisAl 26
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4  AAACCTGAACAGATGCTCCGCACTGATGATGTGGGCATATCGCCTTT 53
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
26  aGlnAlaAlaaspPheMetAlaAsnAspIleThrIleThrGlyLeuGlnA 43
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
54  GGCATTTGCCGACTTCACCATCCCAAGACATCCGTGTGGAAGCTTGCAGC 103
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
43  rgValThrIleGluSerLeuGlnSerValLeuProPheArgLeuGlyGln 59
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
104 GTACCGAGCCGAGCAGCATATTCACACTACCTGCGCCGTCAAAAGTCGGCGAC 153
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
60  ValValSerGluAsnGlnLeuAlaAspGlyValLysAlaLeuTyrAlaTh 76
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
154 ACCTACACAGCACACAGCGGAGTCCCATCATCAAAAGCCTGTACGCCAC 203
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
76  rGlyAsnPheSerAspValGlnValTyrHisGlnGluGlyArgIleIer 93
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
204 CGGTTCCTTCAGCAGCTACGAGTCAAGAACTGCGAGCGGCTGCTTCGCG 253
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
93  yrGlnValThrGluArgProLeuIleAlaGluIleAsnPheGluGlyAsn 109
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
254 TGACCGTTCGTATGCTTCACCTACCATCGGCTCGCTCAACATCACCGGCGCC 303
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
110 ArgLeuIleProLysGluGlyLeuGlnGluGlyLeuLysAsnAlaGlyLe 126
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
304 AAAATGCTGCAAGACGAGCCATCAAGAAACCTCGAATCGTTCGGGCT 353
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
126 uAlaValGlyGlnProLeuLysGlnAlaThrValGlnMetIleGluThrG 143
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
354 GCGCAGTCGCAATACTTAACTAGGCGACACTCAACAGCAGTCGCGCG 403
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
143 LuLeuThrAsnGlnTyrIleSerGlnGlyTyrTyrAsnThrGluIlePhr 159
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
404 GCCTGAAGAAGATATCTCGGCGCGCAAACTCAATATCCAATACAG 453
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
160 ValLysGlnThrMetLeuAspGlyAsnArgValLysLeuAspMetThrPh 176
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
454 CCCAAAGTAACCAACTCGCCGCAACCGCTCGACATCGACATCAGCAT 503
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
176 eAlaGluGlyProAlaArgValValAspIleAsnIleIleGlyAsnG 193
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
504 TGACGAGGGCAAAATCCGCCAAATACCGACATCGAATTTGAAGCAACC 553
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
193 InHisPheSerAspAlaAspLeuIleAspValLeuAlaIleLysAspAsn 209
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
554 AAGTCTATTCGCGCGCAAACTGATCGGCGAGATCTCGCTGACCGAAGGC 603
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
210 LysIle...AsnProLeuSerLysAlaAspArgTyrThrGlnGluLysLe 225
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
604 GGCATTTGGACATGGCTGACACGAAGCGCGGTTTCGACCGCCAGAAAT 653
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
225 uValThrSerLeuGluAsnLeuArgAlaLysTyrLeuAsnAlaGlyPheV 242
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
654 GCGCCAGACATGGAAGAAGTAACCGACTTACCAGAAACACCGCTACT 703
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
242 alArgPheGluIleLysAspAlaLysLeuAsnIleAsnGluAspLysAsn 258
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
704 TCGATTTCCGATCTCGATACCGACATCCAAACCAACGAAGACAAACACC 753
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
259 ArgIlePheValGluIleSerLeuHisGluGlyGluGlnTyrArgPheG1 275
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
754 AGGCAGACCATCAAAATCACCGTCCACGAAGCGGACGTTTCCGCTGGGG 803
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
275 yGlnThrGlnPheLeuGlyAsnLeuThr...TyrThrGlnAlaGluLeuG 291
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
804 CAAGTGTGCGATTAAGCGGACACCAAGAGTCCCCCAAGCGCGCAACTGG 853
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
291 luAlaLeuLeuLysPheLysAlaGluGluGlyPheSerGlnAlaMetLeu 307
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
854 AAAAAGTCTGACCATGAAGCCCGGCAATGTTACGAACGCCAGCAGATG 903
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

308  GluGlnThrThrAsnAsnIleSerThrLysPheGlyAspAspGlyTyrTy 324
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
904 ACCGCGGTTTGGGTGAGATTCAAGAACCGCATGGCTCGGCAGGCTACGC 953
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
324 rTyrAlaGlnIleArgProValThrArgIleAsnAspGluSerArgThrV 341
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
954 ATACAGCGAAATACAGCTACAGCCGCTCGCGAACCGCGGAACCAACCG 1003
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
341 alAspValGluTyrTyrIleAspProValHisProValTyrValArgArg 357
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1004 TCGATTTTCGCTCTGCACATCGAACCGGCGAGAAAATCTACGTCAACGAA 1053
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
358 IleAsnPheThrGlyAsnPheLysThrGlnAspGluValLeuArgArg1 374
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1054 ATCCACATCACCGGCAACAAACAAACCCGCGAGTCTGTCGCGCGGA 1103
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
374 uMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnLeuS 391
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1104 ATTGGCCAAATGGAATCGCGGCTTACGACACCTCCAAAGCTGCAACGCT 1153
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
391 erArgAlaArgLeuMetArgThrGlyPhePheLysHisValThrValAsp 407
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1154 CCAAGAGCGGCTCGAGCTTTGGGCTACTTCGACACGCTACAGTTTGAT 1203
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
408 ThrArgProValProAsnSerProAspGlnValAspValAsnPheValVa 424
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1204 GCGGTCGCGCTTGGCGGTACGCCGCAAAAGTCGATTTGAACATGAGCT 1253
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
424 lGluGlnProSerGlySerSerThrIleAlaAlaGlyTyrSerGlnS 441
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1254 GACGGAAGCTTCCACCGGCTCGCTGACTTGAGCGCGGCTGGGTTCAGG 1303
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
441 erGlyGlyValThrPheGlnPheAspValSerGlnAsnAsnPheMetGly 457
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1304 ATACCGGCTTGGTCACTCGCGCGCGTATCGCAGGCAACCTGTTGCGT 1353
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
458 ThrGlyLysHisValAsnAlaSerPheSerArgSerGluThrArgGluVa 474
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1354 ACGGCAAGTCGCGCGCTCGCGGCTCGCGAAGCAAAACACACGCTCAA 1403
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
474 lTyrSerLeuGlyMetThrAsnProTyrPheThrValAsnGlyValSerG 491
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1404 CGGCTCGCTGCTTTACCGCTACTTTCACGCGACAGCGGGGTACG. 1452
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
491 InSerLeuSerGlyTyr...TyrArgLysThr...LysTyr 502
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1453 .....CTGGGCTACGATATTTACGGAAGCTTCGACCCGCGCAAA 1494
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
503 AspAsnLysAsnIleSerAsnTyrValLeuAspSerTyrGlyGlySerLe 519
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1495 GCATCGACCGCGTCAAAACAATATAAAACCAACACCGCGCGCGGCGT 1544
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
519 userTyrGlyTyrProIleAspGluAsnGlnArgIleSerPheGlyLeuA 536
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1545 AAGGATGGGTATCCCGGTTACGGAATACACCGCGTCAATTCGGGGCTGG 1594
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
536 snAlaAspAsnThrLysLeuHisGlyGlyArgPheMetGlyIleSerAsn 552
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1595 CGCGGAACACCTGACCGTCAAC.....ACCTACAACAAA 1629
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
553 ValLysGlnLeuMetAlaAspGlyGlyLysIleGlnValAspAsnG1 569
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1630 GCACCCAAACGCTATCGCGACTTTATCAACAAATACGGCAAAACCGACGG 1679
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
569 yIleProAspPheLysHisAspTyrThrThrTyrAsnAlaIleLeuGlyT 586
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1680 C...GCACGCGCAGCTTCAAGGCGCTGCTACAAAGGCGACTGTCGCT 1726
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
586 rPAsnTyrSerSerLeuAspArgProValPheProThrGlnGlyMetSer 602
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1727 GGGGCGCAACAAGACCGACGCGCTTATGGCGACGCGCGCTACCTG 1776
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

603  HisSerValAspLeuThrValGlyPhe...GlyAspLys...ThrHisG1 617
```

85-Kilodalton Outer Membrane Protein from *Haemophilus ducreyi*

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525











Brandon,R.C., Fine,L.D., Fritchman,J.L., Fuhrmann,J.L.,  
Georgagen,N.S., Gnehm,C.L., McDonald,L.A., Small,K.V., Fraser,C.M.,  
Smith,H.O. and Venter,J.C.  
Whole-genome random sequencing and assembly of Haemophilus  
influenzae Rd  
Science 269 (5223), 496-512 (1995)  
2 (bases 1 to 10151)  
Taturov,R.L., Mushagian,A.R., Bork,P., Brown,N.P., Hayes,W.S.,  
Borodovsky,M., Rudd,K.E. and Koonin,E.V.  
Metabolism and evolution of Haemophilus influenzae deduced from a  
whole-genome comparison with Escherichia coli  
Curr. Biol. 6 (3), 279-291 (1996)  
3 (bases 1 to 10151)  
White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D.  
Direct Submission  
Submitted (25-JUL-1995) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
4 (bases 1 to 10151)  
White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D.  
Direct Submission  
Submitted (27-SEP-1997) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
The H. influenzae sequence has been updated by R. Fleischmann. New  
database matches have been assigned, product names have been  
improved, and a number of frame shifts have been corrected. We  
gratefully acknowledge the work of Taturov et. al. We have  
incorporated their annotation into the /notes fields of the  
corresponding H. influenzae genes  
5 (bases 1 to 10151)  
White,O., Clayton,R.A., Kerlavage,A.R., Fleischmann,R.D.,  
Peterson,J., Hickey,E., Dodson,R. and Gwinn,M.  
Direct Submission  
Submitted (28-MAY-1998) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
The whole genome was shifted by 588 nucleotides for a new start  
On Sep 30, 1996 this sequence version replaced gi:1221647.  
Location/Qualifiers  
1. .10151  
/organism="Haemophilus influenzae Rd"  
/db\_xref="taxon:71421"  
81. .845  
/gene="HI0912"  
81. .845  
/gene="HI0912"  
/note="hypothetical protein; identified by GeneMark;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="H. influenzae predicted coding region HI0912"  
/protein\_id="AAC22579.1"  
/db\_xref="GI:1573933"  
/translation="MNNKISVYDKDNFFELYQKLRANPISLNEIIEKPTWLSLLPNLK  
GKLLDLCGCTGHQLYLRGAQKAVIGTDLSEKMLEQAEKDLQKCGQFSGRFSLYHL  
PIEKLAEPSHFDVITSSFAFYIENFTLSEKMLEQAEKDLQKCGQFSGRFSLYHL  
EGERWEKNDKQOVAYRLNHYREGRKNRWKQFPQYHRTTATINLIHARFQIE  
QMEPEMLADQPOWHNEFKDLSHRPPILLFIKARKVEK"  
990. .1745  
/gene="HI0913"  
990. .1745  
/gene="HI0913"  
/note="similar to GB:D26562 SP:P02351 GB:V00343 PID:42842  
PID:473824 percent identity: 82.16; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="ribosomal protein S2 (rps2)"  
/protein\_id="AAC22571.1"  
/db\_xref="GI:1573934"  
/translation="MEAPQOKENIMAOVSMRDMINAGVHFHGTQRYNPMQPFPIFG  
ARNGVHINLEKTLPLFNEALIELTRASNNGKVLVFGTKRAAEAVQAALQOQYV  
VNHWRGGMLTNKVTNQSIKRLKDLKETQSQDQFTFKLTKKEALMRSEKLELSLG

GIKMGGLPDALFVIGADHEHIAVKEANNLIGIPVFAIVDTNSTPAGVDFVPIPGNDAT  
RAIQLYSAASAAAVKEGRNEAQVAEELAADAEE"  
1878. .2729  
/gene="HI0914"  
1878. .2729  
/gene="HI0914"  
/note="similar to GB:D26562 SP:P02997 GB:D13334 GB:V00343  
PID:1208943 percent identity: 71.43; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="elongation factor Ts (tsf)"  
/protein\_id="AAC22572.1"  
/db\_xref="GI:1573935"  
/translation="MAETASLVKELRDTGAGMMECKKALVEANGDIELAIDNMRKS  
GQAKAKKAGRAEAGVILARVENGFGVLYEMNCETDFVAKDAGFLGIANEYTFDAAA  
NKGTTTIALQAQFEERKALVAKIGENNRIRVAYLDGQVIAIQLHGAKIGLVVAGEG  
SADELKAMHVAASKPEFVNPEDVSAEVEHEHQIQIDIAINSGKPEKAEKRVGEGR  
MKRPTGEVSLTGQAFVMDPSVSGDFLKSVTNSVSNFIRLEVGEKIEKEDFAAEVA  
KITGGNA"  
complement(2909. .3934)  
/gene="HI0915"  
complement(2909. .3934)  
/gene="HI0915"  
/note="similar to GB:X54797 SP:P21645 PID:1208950  
PID:41470 GB:U00096 percent identity: 65.58; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="UDP-3-O- (3-hydroxymyristoyl)-glucosamine  
N-acetyltransferase (lpxB)"  
/protein\_id="AAC22573.1"  
/db\_xref="GI:1573936"  
/translation="MQKSYSLQELATOIGATVRCNTDVVNTAPLDKASQNSLTFLS  
NWFRLKDSKAGILVSEEDVEHCSPESNLLIVKDPYVAYAILAQIMYSTPKAAAG  
IAKSAIFDGVGLNGISGANAVIEGVVLDGNDVIIGANCVFGKNTKIGSTOLWAN  
VTYVHNVEIGANLLIQSGTVIGSDFGFYANDRGWRKIPQVGVVIGNNVEIGANTCI  
DGRALDATTIEDNVIDNLQIAHNHVGITGAVAGVIMAGSLTVGRYCLIGASVI  
NGHMEICDQVITITGMGMVRCPIPEPGVYSSGIPLOTNKEWRKTAALTGLDIGNKRLK  
ALEKKIS"  
complement(3947. .4540)  
/gene="HI0916"  
complement(3947. .4540)  
/gene="HI0916"  
/note="similar to PID:1935023 percent identity: 37.58;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="outer membrane protein, putative"  
/protein\_id="AAC22574.1"  
/db\_xref="GI:1573937"  
/translation="MKNIAKVTAALGIALASGYASAEKIAFINAGYIFQHPHPRQA  
VADKLDAEFKPAEKLAASKVEDDKTAAARKKVEAKVALEKADAPRLROADIQKROO  
EINKLGAEDAEQLKMQEDKKVQFOAQNEKRAEKERKLLDSIQATNNLAKAG  
YTVLDANSIVFAVEGKDITEEVLKUSIPASEKAQEKK"  
complement(4648. .7035)  
/gene="HI0917"  
complement(4648. .7035)  
/gene="HI0917"  
/note="similar to GB:U13961 PID:1573938 PID:2231192  
PID:2231195 percent identity: 98.74; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="protective surface antigen D15"  
/protein\_id="AAC22575.1"  
/db\_xref="GI:1573938"  
/translation="MKLLJASLLFGTTTTFVAFPAFVAKDIRVDGVGDLEBQQLIRASL  
IPVARGQRTDNDVANIIVRSFLVSGRFDVKAHQEGDVLVSVVAKSIISDVKIGKNSI  
IPTPEALKONLDANGFKVGDVILIREKLNFAKSVKHEVSVGRYNAIVPIVNTLPNNR  
AEILQIYNEDDKAKLASLTFKGNESSVSSLTQEQMLQDPSWKLWNKFGPAQPEKD  
LOSTRDYLANGYAKAQTITQDVLNDEKTKVNTYIDNSELQVLDLSARLIIGLGM  
SABLEPLLSALHLNDFRRSDIADVENAIKAKLGERGYGSAVTNVSVPDFDANKTLAI









```
516 GlyGlySerLeuSerTyrGlyTyrProLeuAspGluAsnGlnArgIleSe 532
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1690 CGAAGTAATGTTACTTTAGGTTTCCCTTAATGAAATAAATCAATCTATTA 1739
532 rPheGlyLeu.....AsnAlaAspAsnThrLysLeuHisG 544
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1740 TGTAGGATAGGTGCATACCTATAATAAATAGTAACTTGTCTAGAAAT 1789
544 lyGlyArgPheMetGlyIleSerAsnValLysGlnLeuMetAlaAspGly 560
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1790 ATAACCGTAATTTATATATCAATCAATGAAA..... 1821
561 GlyLysIleGlnValAspAsnAsnGlyIleProAspPheLysHisAspTy 577
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1822 .....TTTAAAGGTAATGCCATTAACAAATGACTTTGATT 1859
577 rThrThrTyrAsnAlaIleLeuGlyTyrAsnTyrSerSerLeuAspArgp 594
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1860 TTCT.....TTTGGTTGGAAGTATACACAGCCTTAATAGAG 1894
594 roValPheProThrGlnGlyMetSerHisSerValAsp.....LeuThr 608
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1895 GCTATTTCCCACTAAAGGGTTAAGCAAGTCTTGGTGGACGAGTTACT 1944
609 ValGlyPheGlyAspLysThrHisGlnLysValValTyrGlnGlyAsnII 625
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1945 ATTCCAGGTTCTGATAACAAATACTACAACTAAAGTGCAGATGTACAGG 1994
625 eTyrArgProPhe.....IleLysLysSerValI 635
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1995 TTCTTACCATTAGACAGAGATCACCTCTGGTTGTATCTGCAAAAGCAT 2044
635 euArgGlyTyrAlaLysLeuGlyTyrGlyAsn...AsnLeuProPheTyr 650
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2045 CTGCAGGATATGCAAT...GGTTTGGAAACAACGGTTTACCGTTCTAT 2091
651 GluAsnPheTyrAlaGlyClyTyrGlySerValArgGlyTyrAspGlnSe 667
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2092 CAAACTTATACAGCGGGTGGCATCGTTCATTACGTGGTTTGTCTTATGG 2141
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgGlyGlnG 684
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2142 TAGATTGGACCTACGCA.....ATTATGCCGAATATGTAATG 2182
684 InThrThrLeuGly.....GluValValGlyGly 693
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2183 GTAGTGGTACTGGTACTTTTAAAGAGATAAGTCTCTGATGTGTTGGTGT 2232
694 AsnAlaLeuAlaThrPheGlySerGluLeuLeuLeuProLeuProPheTy 710
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2233 AATGCAATCGCTACAGCTAGCGCAGAGTAAATGTGCCAACTCCATTGT 2282
710 sGlyAspTrpIle...AspGlnValArgProValIlePheIleGluGly 726
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2283 GAGCGATAAGACCCAAATACGGTCCGAACGCTCTTATTGTGTATGCGG 2332
726 lyGlnValPheAspThrThr...GlyMetAspLysGlnThrIleAspLeu 741
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2333 CAAGTGTGTGGAATCACTAAATGGAATCAGATAAAATGGATTAGAGAGC 2382
742 ThrGlnPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaI 758
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2383 GATGTATTAAAA..... 2394
758 aAsnArgProLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyV 775
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2395 .AGATTGCTCTGATTATGCAATCAAGCCGATTTCGCCCTCTACAGGTG 2443
775 alGlyAlaThrTyrThrProIleGlyProLeuSerIleSerTyrAla 791
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2444 TCGGATTCATGCGCAATCTCTATTGGCCATTGGTATTCTCTATGCC 2493
792 LysProLeuAsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGl 808
```

```
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2494 AAACCAATATAAAATATCAAAATGATGATCGAACAGCTTCCAATTAG 2543
808 nIleGlySerValPhe 813
|||||:|||||:|||||:|||||:|||||:|||||:
2544 TATTGGAGGTTCTTTC 2559
seq_name: gb_ba:HIU60833
seq_documentation_block:
LOCUS HIU60833 2776 bp DNA linear BCT 02-JUL-1997
DEFINITION Haemophilus influenzae D15 (d15) gene, complete cds.
ACCESSION U60833
VERSION U60833.1 GI:2231194
KEYWORDS
SOURCE
ORGANISM
Haemophilus influenzae.
Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
Haemophilus.
REFERENCE
1 (bases 1 to 2776)
AUTHORS
Loosmore,S.M., Yang,Y., Coleman,D.C., Shortreed,J.M., England,D.M.
and Klein,M.H.
TITLE
The gene encoding outer membrane protein D15 is highly conserved
amongst Haemophilus influenzae species and may be part of an operon
Unpublished
REFERENCE
2 (bases 1 to 2776)
AUTHORS
Loosmore,S.M.
TITLE
Direct Submission
JOURNAL
Submitted (13-JUN-1996) Loosmore S. M., Connaught Laboratories
Limited, Molecular Genetics Research, 1755 Steeles Avenue W., North
York, Ontario M2R 3T4, Canada
FEATURES
source
1..2776
/organism="Haemophilus influenzae"
/isolate="Minna"
/species="Homo sapiens"
/db_xref="taxon:727"
/clone="DS-691-1-5"
328..2721
/gene="d15"
328..2721
/gene="d15"
/codon_start=1
/transl_table=11
/product="D15"
/protein_id="AAB61976.1"
/db_xref="GI:2231195"
/translation="MKLLIASLLFGTTTTFVFAAPVFAKDIRVDGVQGLDLEQIRASI,
PVRAGQRTVDNDVANIVRSIFVSGRFDDVKAHQEDVLVSVVAKSIISDVKIRGNSV
IPTALKQNDLAKNGFKVGDVLIREKLNFEAKSVKHEHYASVGRYNATVEPTVNTLPNNR
AEILIOINEDDKAKLASLTFKGNESVSSSTLOEOMELQPDSSWKLGNKFGAOFKED
LOSIRDYILNNGYAKAQTITKTDVQNDKTKVNTIDVNEGLOYDLRSARIIGNLGM
SAELPILLSALHLNDTFRSDIADVENAIKALGERGYSATVNSYPDFDDANKTLAI
TLVDAGRLTLVRLRFEPTVSADSLRQEMRQEGTWTNSQLVELGKIRLDTGFF
ETVENRIDPINGSNDVYVYKVKERTGSINFGIGYTESGSIYQASVKODNPLCTG
AAVSIAGTKNDYTSVNLGYTEPTFKDGVSLGNGVFFENYDNKSDTSNRYKRTYTG
SNVTLPFPVNNNSYVGLGHTYKINSFALYRNLYIQSMFKGNGIKTNDDFSF
GWYNSLNRGYFPTKGVKASLGRVTPGSDNKKYKLSADVQGFYPLDRDLHWVSAS
ASAGYANGFGNKLPPFYQYTAGTIGSLRFAYSIGPNALYAEYNGSGTCTGFKKIS
SDVIGGNATASAEILVPTFPVSDSKSNTVTSLSFVDAASVWNTWKSDKNGLSDV
LKRLPDYKGKSSRIRASTGFGWFOGSPIGLTVFYAKPIKKYENDVDEQFSGGSP"
BASE COUNT 888 a 435 c 588 g 865 t
ORIGIN
```

```
alignment_scores:
Quality: 1052.50 Length: 822
Ratio: 1.956 Gaps: 20
Percent Similarity: 65.450 Percent Identity: 31.630
alignment_block:
US-09-701-711-2 x HIU60833
```





[illegible]

419	AspValAsnPheValValGluGluGlnProSerGlySerSerThrIleAl	435
1617	GATGCTGATATAAAGCTAAAGACGTAACACGGGTAGTATCAACTTTGG	1666
435	aAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSerG	452
1667	TATTGGTTACGTCACAGAGTGGTATCAGTTATCAACAAGATATTAAC	1716
452	InAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerArg	468
1717	AAGATAATTTCTTGGGAACAGGGCGCGCATAGTAGCTGGTACGAAA	1766
469	SerGlnThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheTh	485
1767	ARTGATTATGGTACGAGTGCATTTGGGTATACCGAACCCCTATTATC	1816
485	rValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLysT	502
1817	TAAAGATGGTGTA....AGTCTTGGTGGAAATATTTCTTTGAAAACT	1860
502	yrAspAsnLys.....AsnIleSerAsnTyrValLeuAspSerTyr	515
1861	ACGATAACTCTAAAGTGAACATCCCTCACTATATAGCGTCAGACTTAT	1910
516	GlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSe	532
1911	GGAAGTAATGTTACTTTAGGTTCCCTGTAATCAAAATAAATCCCTATTA	1960
532	rPheGlyLeu.....AsnAlaAspAsnThrLysLeuHisG	544
1961	TGTAGGATATAGGCCATACCTATATAAAATAGTAATTTGCTCTAGAAT	2010
544	LyGlyArgPheMetGlyIleSerAsnValLysGlnLeuMetAlaAspGly	560
2011	ATAACCGTAATTTATATATCAATCAATGAAA.....	2042
561	GlyLysIleGlnValAspAsnAsnGlyIleProAspPheLysHisAspTy	577
2043	.....TTTAAAGGTAATGGCATTTAAACAATAACACTTTCATTT	2080
577	rThrTyrAsnAlaIleLeuGlyTyrAsnTyrSerSerLeuAspArgp	594
2081	TTCT.....TTTGGTGGCACTATACAGCGCTTAATAGAG	2115
594	roValPheProThrGlnGlyMetSerHisSerValAsp.....LeuThr	608
2116	GCTATTTCCCAACTAAAGGGGTTAAAGCAAGCTCTGGTGGACGAGTTACT	2165
609	ValGlyPheGlyAspLysThrHisGlnLysValVal.....TyrGlnGI	623
2166	ATTCAGGTTCTGTGATACAAATACTACAAACTAAGTCAGATGTACAGGG	2215
623	y.....AsnIleTyrArgProPheIleLysLysSerValI	635
2216	TTTTCTACCATTAGACAGAGATCACCGCTGGTGTATCTGCAAAAGCAT	2265
635	euArgGlyTyrAlaLysLeuGlyTyrGlyAsn...AsnLeuProPheTyr	650
2266	CTGCAGAGATATGCAAAAT...GGTtTTGGAAACAAGGGTTTTACCGTTCAT	2312
651	GluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe	667
2313	CAAACTTATACAGGGGTGCATTTGGTTCATACGGCGGTTTTGCTTATGG	2362
667	rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgGlyGlyGlnG	684
2363	TAGTATTGGCCCTAATGCA.....ATTATGCCGAACATGCTAATC	2403
684	lnThr.....ThrLeuGlyGluValValGlyGlyAsnAlaLeuAla	697
2404	GTACTTTTAAGATAAGTTCGTGATGCTATGGTGTATGGTGAATGCAATCACA	2453









DEFINITION Sequence 1 from patent US 6083743.

ACCESSION AR102148

VERSION AR102148.1 GI:12812946

KEYWORDS

SOURCE

ORGANISM

Unknown.

Unclassified.

REFERENCE 1 (bases 1 to 2949)

AUTHORS Chong, P., Thomas, W., Yang, Y. Ping, Loosmore, S., Sia, D. Yuan Charles

TITLE Haemophilus outer membrane protein

JOURNAL Patent: US 6083743-A 1 04-JUL-2000;

FEATURES Location/Qualifiers

1..2949

source /organism="unknown"

BASE COUNT 984 a 475 c 613 g 877 t

ORIGIN

alignment\_scores:

Quality: 1052.50 Length: 822

Ratio: 1.956 Gaps: 20

Percent Similarity: 65.450 Percent Identity: 31.630

alignment\_block:

US-09-701-711-2 x AR102148 ..

Align seg 1/1 to: AR102148 from: 1 to: 2949

23 SerThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIleThrIleTh 39

114 ACAAGCTGTTGGCGACCTTTGTGGCAAAAGATATTCGTGGGA 163

39 rclyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPheA 56

164 TGGTGTCAAGGTGACTTAGAACAAATCCGACGAGTTTACCTGTTTC 213

56 rgLeuGlyGlnValValSerGluAsnGlnLeuAlaAspGlyValLysAla 72

214 GTGCGGTGACGCGTGCTGACTGACAAATGATGTGGCTAATATTGTCGCTCT 263

73 LeuTyAlaThrGlyAsnPheSerAspValGlnValTyHisGlnGluG1 89

264 TTATTCGTAAGTGTGCTGATTCGATGATGTGAAGCG...CATCAAGAGG 310

89 yArgIle...IleTyGlnValThrGluArgProLeuIleAlaGluLea 105

311 CGATGTGCTTGTGTAGCGTTGTGGCTAAATCGATCATTTTCAGATGTTA 360

105 snPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluGlyLeu 121

361 AAATCAAGGTAACCTCTGTTATTCCTCACTGAAGCACTTAAACAAACTTA 410

122 LysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrValG1 138

411 GATCCTAACGGTTTAAAGTTGGCGATGTTTAAATTCGAGAAATTA 460

138 nMetileGluThrGluLeuThrAsnGlnTyIleSerGlnGlyTyTyra 155

461 TGAATTTGCCAAAAGTGTAAAGAGCACTATGCAAGTGTAGTGCCTATA 510

155 snThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgValLys 171

511 ACGCAACAGTTGAACCTATTGTCATAACGCTACCAAAATATTCGCGCTGA 560

172 LeuAspMetThrPheAlaGluGlyLysProAlaArgValValAspIleAs 188

561 ATTTTATTCAAATCATGATGATGATGATGATGATGATGATGATGATGAT 610

188 nIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspValLeuA 205

611 TTTCAGGGGAACGAATCTGTGTAGTACGATGATGATGATGATGATGATGAT 660

205 laileLysAspAsnLysIleAsnProLeuSerLysAlaAspArgTyThr 221

661 AATTACACCTGATTCCTGGTGAATA...TGGGGAATAAATTGAA 707

222 GlnGluLysLeuValThrSerLeuGluAsnLeuArgAlaLysTyLeuAs 238

708 GGTGGCAATTCGAGAAAGATTTCAGTCAATTCGTGATTATTATTAA 757

238 nAlaGlyPheValArgPheGluIleLysAspAlaLysLeuAsnIleAsnG 255

758 TAATGGCTATGCCAAAGACAAATTAATAACGGATGTTTCAGCTAAATG 807

255 luAspLysAsnArgIlePheValGluIleSerLeuHisGluGlyGluGln 271

808 ATGAAAAACAAGATTAAATGTAACCATTTGATGATAAATGAAGTTTACAG 857

272 TyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuThrTyThrGlnAl 288

858 TATGACCTTCGTAGTGCACGATTAATAGTAATCTGGGAGGTATGCTGC 907

288 aGluLeuGluAlaLeuLys...PheLysAlaGluGluGlyPheSerG 304

908 CGAGCTTGACCTTTACTTTTCAGCATTTACATTTAAATGATACTTCCGCC 957

304 InAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAsp 320

958 GTAGTGATATTGCAGATGTAGAAAAATGCAATTAAGCAAAACTTGAGAA 1007

321 AspGlyTyTyTyAlaGlnIleArgProValThrArgIleAsnAspG1 337

1008 CGCGTTACGGTAGCGCAACGGTAAATTCAGTACCTGATTTTGTGATGTC 1057

337 uSerArgThrValAspValGluTyTyIleAspProValHisProValT 354

1058 AAATAAACAATTCAGGATAACCTTGTGTGATGCTGGACGACGTTAA 1107

354 yrValArgArgIleAsnPheThrGlyAsnPheLysThrGlnAspGluVal 370

1108 CTGTTCGCAACTTCGCTTTGAAGAAATACCGTTTCTGCTGATAGCACT 1157

371 LeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnly 387

1158 TTACGTCAGGAATTCGCCAACAGAGAACTTGGTATATATTCACAAT 1207

387 sileGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePheLysHisV 404

1208 AGTTGAGTTAGGAAAAATTCCTTAGATCGTACAGGTTTCTCGAA... 1253

404 alThrValAspThrArg.....ProValProAsnSerProAspGlnVal 418

1254 ..ACAGTCGAAACCGAATTCATCTATCAATGGTAGTAAATGATGAAGTG 1301

419 AspValAsnPheValValGluGluGlnProSerGlySerSerThrIleAl 435

1302 GATGTCGTATATAAGTCAAGAACGTAACAGGCTAGTATCAACTTTGG 1351

435 aAlaGlyTyTySerGlnSerGlyGlyValThrPheGlnPheAspValSerG 452

1352 TATTGGTTACGCTACAGAGAGTGGTATTAGTTATCAAGCAAGTGTAAAC 1401

452 InAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468

1402 AAGATAATTTCTGGGAACGGGCGCAGTAAGTATAGCTGGTACGAAA 1451

469 SerGluThrArgGluValTyTySerLeuGlyMetThrAsnProTyThrPheTh 485

1452 ATGATTTATGTCAGTGTCAATTTGGTTATACCGAGCCCTATTATTAC 1501

485 nValAsnGlyValSerGlnSerLeuSerGlyTyTyTyArgLysThrLysT 502

1502 TAAAGATGGTGTGA...AGTCTTGGTGGAAATGTTTCTTCTTGAAGAACT 1545

502 yrAspAsnLys.....AsnIleSerAsnTyTyValLeuAspSerTy 515

```
|||||...::: |||||...::: |||||...::: |||||...::: |||||...:::
1546 ACATAACTCTAAAGATGATACATCTCTAACTATAAGCGTACGACTTAC
516 GlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSe 532
|||||...::: |||||...::: |||||...::: |||||...::: |||||...:::
1596 GGAAGTAATGTTACTTAGTTCCCTGTAATGAAATAGAAATACCTCTATTA 1645
532 rPheGlyLeu...::: AsnAlaAspAsnThrLysLeuHisG 544
|||||...::: |||||...::: |||||...::: |||||...::: |||||...:::
1646 TGTAGGATTAGTTCATACCTATAATAAAATAGTAAGTTCCTCTAGAAT 1695
544 lyGlyArgPheMetGlyIleSerAsnValLysGlnLeuMetAlaAspGly 560
|||||...::: |||||...::: |||||...::: |||||...::: |||||...:::
1696 ATAACCGTAATATATATATCAATCAATGAAA...::: 1727
561 GlyLysIleGlnValAspAsnAsnGlyIleProAspPheLysHisAspTy 577
|||||...::: |||||...::: |||||...::: |||||...::: |||||...:::
1728 .....TTTAAGGTAATGCAATTAACAAATACACTTTGATTT 1765
577 rThrThrTyrAsnAlaIleLeuGlyTrpAsnTyrSerSerLeuAspArgP 594
|||||...::: |||||...::: |||||...::: |||||...::: |||||...:::
1766 TTCT.....TTTGGTGAACATATAACAGCCTTAATAGAG 1800
594 roValPheProThrGlnGlyMetSerHisSerValasp.....LeuThr 608
|||||...::: |||||...::: |||||...::: |||||...::: |||||...:::
1801 GCATTTTCCCACTAAAGGGGTTAAAGCAAGTCTTGGTGACGAGTTACT 1850
609 ValGlyPheGlyAspLysThrHisGlnLysValValTyrGlnGlyAsnII 625
|||||...::: |||||...::: |||||...::: |||||...::: |||||...:::
1851 ATTCCAGGTTCTGATACAAATACTACAACTAAGTGCAAGTATGACAGG 1900
625 etYrArgProPhe.....IleLysLysSerValL 635
|||||...::: |||||...::: |||||...::: |||||...::: |||||...:::
1901 TTCTACCCATTAGACAGAGATCACCTCGGTGTATCTGCAAAAGCAT 1950
635 euArgGlyTyrAlaLysLeuGlyTyrGlyAsn...AsnLeuProPheTyr 650
|||||...::: |||||...::: |||||...::: |||||...::: |||||...:::
1951 CTCAGGATATGCAAA...GGTTTGGAAACAAGCGTTTACCGTCTAT 1997
651 GluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667
|||||...::: |||||...::: |||||...::: |||||...::: |||||...:::
1998 CAACTTATACAGCGGTGCATCGGTTCATTACGCGTGTTCCTATGG 2047
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgGlyGlnG 684
|||||...::: |||||...::: |||||...::: |||||...::: |||||...:::
2048 TAGTATTGGACCTAACGCA.....ATTATGCGGAATATGGTAATG 2088
684 lnThrThrLeuGly.....GluValValGlyGly 693
|||||...::: |||||...::: |||||...::: |||||...::: |||||...:::
2089 GTAGTGGTACTGTGTTTAAAGAGATAAGTTCTGTGTTGTTGGT 2138
694 AspAlaLeuAlaThrPheGlySerGluLeuIleLeuProLeuProPheLy 710
|||||...::: |||||...::: |||||...::: |||||...::: |||||...:::
2139 AATGCAATCGCTACAGCTACGCGAGAGTAATTTGTCCTCACTCCATTGT 2188
710 sGlyAspTrpIle...AspGlnValArgProValIlePheIleGluGly 726
|||||...::: |||||...::: |||||...::: |||||...::: |||||...:::
2189 GAGCGATAAGAGCGAAATACGTCGGAACCTCTCTTATTGTTGATGCGG 2238
726 lyGlnValPheAspThrThr...GlyMetAspLysGlnThrIleAspLeu 741
|||||...::: |||||...::: |||||...::: |||||...::: |||||...:::
2239 CAAGTCTTTGGAACTAATAATGAAATCAGATAAAATGGATTAGAGAGC 2288
742 ThrGlnPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaI 758
|||||...::: |||||...::: |||||...::: |||||...::: |||||...:::
2289 GATGTAATAA..... 2300
758 aAsnArgProLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGly 775
|||||...::: |||||...::: |||||...::: |||||...::: |||||...:::
2301 .AGATTGCTGATTATGGAATAACGCGTATTCGCGCTCTACAGGTG 2349
775 aGlyAlaThrTrpTyrThrProIleGlyProLeuSerIleSerTyrAla 791
|||||...::: |||||...::: |||||...::: |||||...::: |||||...:::
```

```
2350 TCGGATTCCAATGGCAATCTCTATTGGGCAATTGGTATTCTTCTATGCC 2399
792 LysProLeuAsnLysLysGlnAsnAspGlnThrAspThrValGlnPheG1 808
|||||...::: |||||...::: |||||...::: |||||...::: |||||...:::
2400 AAACCAATTAATAAATGAAATGATGATGTCGACACAGTTCCANTTAG 2449
808 nileGlySerValPhe 813
|||||...::: |||||...::: |||||...::: |||||...::: |||||...:::
2450 TATTGGAGGTTCTTTC 2465
seq_name: gb_pat:A38694
seq_documentation_block:
LOCUS A38694
DEFINITION Sequence 5 from Patent WO9412641.
ACCESSION A38694
VERSION A38694.1 GI:2295177
KEYWORDS
SOURCE
ORGANISM
unidentified.
unclassified.
REFERENCE
1 (bases 1 to 2950)
AUTHORS Chong,P., Thomas,W., Yang,Y., Loosmore,S., Sia,D.Y. and Klein,M.
TITLE HAEMOPHILUS OUTER MEMBRANE PROTEIN
JOURNAL Patent: WO 9412641-A 5 09-JUN-1994;
CONNAUGHT LAB (CA)
COMMENT Other publication CA 2149319 940609
Other publication AU 5556594 940622
Other publication JP 8502417T 960319.
FEATURES
Location/Qualifiers
1..2950
/organism="unidentified"
/db_xref="taxon:32644"
334..2727
/note="unnamed protein product"
/codon_start=1
/protein_id="CAA02390.1"
/db_xref="GI:2295178"
/translation="MKLLIASLFLGTTTTFVFAAPFAKDIRVDGVQGLDLEQIRASL
PVRAQRTDNDVANIIVSRSLFVSGRFDDVKAHQEGDVLVSVVAKSIIISDVLIKNSV
IPTALKQNDLDRANGFKVGLIRKLENEFAKSVKHEIYASVGRYNATVEIVTLPNNR
AEILQINEDDKAKLASLTFKGNESVSSSTLOEQMELQDPSWMLGWKFEQAQFEKD
LQSIDRYLLNNGYAKAQTKTQDLNDEKTYNVTIDVNEGLOYDLRSARIIGNLGM
SALEPLLSALHLNDFRSDIADVENAIKAKLGBRGYSATVNSVPPDDANKTLAI
TLVVDAGRLTVROLRFEGTYSADSTLROEMRQEGTWYNSQLVELGKIRLDRGTGF
ETVENRIDPLNGSNDVDVYKVKERTGSIINFGICGTESGISVQASVKQDNFLGTG
AAVSIAGKNDYGTSVNLGYTEPYFKDGSVSGNFFPENYDNSDTSNTKRTTYG
SNVTILGFPVNNNSYVGLGHTYINKISFALNRYNRLYIQSMKFKGNGIKTNDDFS
GMVYNSLNRGYFPTKGVKASLGGRTVIFGSDNKYKLSADVQGFPLDRHLWVVSAS
ASAGYANGFGNKRLLPFQYTYTAGGIGSLRGFAYSIGPNAIYAENYNGSGTGTFRKIS
SDVIGGNAIATASAEIYTPFPVSDKSONTVTSFLFVDAASVWNTKWKSDKNGLESFV
LKRLPDYCKSSRIRASTGVGFQWQSPIGPLVFSYAKPIKKIYENDDDVEQFQFSIGSF"
BASE COUNT 950 a 465 c 619 g 916 t
ORIGIN
alignment_scores:
Quality: 1052.50 Length: 822
Ratio: 1.956 Gaps: 20
Percent Similarity: 65.450 Percent Identity: 31.630
alignment_block:
US-09-701-711-2 x A38694
Align seg 1/1 to: A38694 from: 1 to: 2950
23 SerThrHisAlaGlnAlaAspPheMetAlaAsnAspIleThrIleTh 39
::: ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
373 ACAACGACTGTGTTGCGCACCTTTTGTGCGCAAAAGATATTCGTGTGA 422
::: ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
39 rGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPhea 56
||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
423 TGGTGTTCAGAGTGACTTAGAACAAACAAATCCGACGAAGTTTACCTGTC 472
```

```
56  rgLeuGlyGlnValValSerGluAsnGlnLeuAlaAspGlyVallysAla 72
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
473 GTGCGGTACCGGTGACTGACATGATGATGGCTAATATGTCGCTCT 522
73  LeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisGlnGlu 89
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
523 TTATTCGTAAGTTCGATTCGATGATGAAAGCG...CATCAAGAGG 569
89  yArgile...lleTyrGlnValThrGluArgProLeuIleAlaGluIleA 105
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
570 CGATGTGCTGTTGTTAGCTGTGGCTAAATCGATCATTTTCAGATGTA 619
105 snPheGluGlyAsnArgIleProLysGluGlyLeuGlnGlnGlyLeu 121
   : : : : : | : : : : | : : : : | : : : : | : : : : |
620 AAATCAAGAGTAACCTCTGTTATTCCTCACTGAAGCCTTAAACAAACTTA 669
122 LysAsnAlaGlyLeuAlaValGlyGlnProLeuGlyGlnAlaThrVal 138
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
670 GATGCTAACGGGTTTAAAGTTGGCGATGTTTAAATTCGAGAAAAATTA 719
138 nMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyrVal 155
   : : : : : | : : : : | : : : : | : : : : | : : : : |
720 TGAATTTCCCAAGAGTGAAGAGCCTATGCAAGTGTAGTGCCTATA 769
155 snThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgVallys 171
   || : : : : : | : : : : | : : : : | : : : : | : : : : |
770 ACGCAACAGTTGAACCTATTGTCATAGCTACCACCAATAATCGCGTCAA 819
172 LeuAspMetThrPheAlaGluGlyLysProAlaArgValValaspIleAs 188
   || : : : : : | : : : : | : : : : | : : : : | : : : : |
820 ATTTTAAATCAAAATCAAGATGATAAGCAAAATTTGGCATTCATTAAC 869
188 nIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleValLeuA 205
   : : : : : | : : : : | : : : : | : : : : | : : : : |
870 TTTCAAGGGGAACCAATCTGTTAGTCAGTACATTAACAAGCAAAATGG 919
205 laIleLysAspAsnLysIleAsnProLeuSerLysAlaAspArgTyrThr 221
   : : : : : | : : : : | : : : : | : : : : | : : : : |
920 AATTACAACTGATCTTGGTGGAAATTA...TGGGGAATAAATTTCAA 966
222 GlnGluLysLeuValThrSerLeuGlnAsnLeuArgAlaLysTyrLeuAs 238
   : : : : : | : : : : | : : : : | : : : : | : : : : |
967 GGTGCGCAATTCGAGAAAGATTTTCAGTCAATTCGTGATTATTATAA 1016
238 nAlaGlyPheValArgPheGluIleLysAspAlaLysLeuAsnIleAsnG 255
   | : : : : : | : : : : | : : : : | : : : : | : : : : |
1017 TAATGGCTATGCCAAGCACAAATTAATAAACCGGATGTTCCAGCTAAATG 1066
255 luAspLysAsnArgIlePheValGluIleSerLeuHisGluGlyGluGln 271
   : : : : : | : : : : | : : : : | : : : : | : : : : |
1067 ATGAAAAACAACAAATTAATCTAACCATTTGATGTAATGAAGGTTACAG 1116
272 TyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAl 288
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1117 TATGACCTTCGTAGTCAGCGCATATAGTAATCTGGGAGGTATGTCGC 1166
288 aGluLeuAlaLeuLeuLys...PheLysAlaGluGluGlyPheSerG 304
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1167 CGAGCTTGAACCTTTTACCTTCAGCATTCATATTAATGATACTTTCCGCC 1216
304 lnAlaMetLeuGluGlnThrThrAsnIleSerThrLysPheGlyAsp 320
   : : : : : | : : : : | : : : : | : : : : | : : : : |
1217 GTAGTGATATTCAGATGTAACAAATGCAATTAAGCAAAACTTTGGAGAA 1266
321 AspGlyTyrTyrTyrAlaGlnIleArgProValThrArgIleAsnAspG 337
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1267 CGCGGTTACGGTAGCGAACCGGTAATTCAGTACCTGATTTTGTATGATGC 1316
337 uSerArgThrValAspValGluTyrTyrIleAspProValHisProValT 354
   : : : : : | : : : : | : : : : | : : : : | : : : : |
1317 AAATAAACATATTCGATAACCCCTGTTGTTGATGCTGGGACGCGTTAA 1366
```

```
354  yrValArgArgIleAsnPheThrGlyAsnPheLysThrGlnAspGluVal 370
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1367 CTGTTCCCAACTTCGCTTTGAAGAAATACCGTTCTCTGCTGATGACACT 1416
371 LeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLys 387
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1417 TTACGTCAGGAATTCGCCCAACAAGAAAGAACTTGGTATATAATTCACAAT 1466
387 sIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePheLysHisV 404
   : : : : : | : : : : | : : : : | : : : : | : : : : |
1467 AGTTGAGTTAGAAAAATTCGCTTAGATCGTACAGGTTCTTCGAA... 1512
404 alThrValAspThrArg...ProValProAsnSerProAspGlnVal 418
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1513 ..ACAGTCGAAACCGAATTCATCTATCAATGGTAGTAAATGAAGCTG 1560
419 AspValAsnPheValValGluGluGlnProSerGlySerSerThrIleAl 435
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1561 GATGTCGTATATAAGTCAAGAACGTAACACGGGTAGTATCAACTTTGG 1610
435 aAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSerG 452
   : : : : : | : : : : | : : : : | : : : : | : : : : |
1611 TATTGGTTACGCTACAGAGAGTGGTATTAGTTATCAACCAACTGTTAAAC 1660
452 lnAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1661 AAGATAATTTCTTGGGAACAGGGCGGCGAGTAAGTATAGCTGGTACGAAA 1710
469 SerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheTh 485
   : : : : : | : : : : | : : : : | : : : : | : : : : |
1711 AATGATTATGGTACGAGTGTCAATTTGGGTTATACCGAGCCCTATTTTAC 1760
485 rValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLysT 502
   | : : : : | : : : : | : : : : | : : : : | : : : : |
1761 TAAAGATGGTGTA...AGTCTTGGTGGAAATGTTTCTTTGAAAACT 1804
502 yrAspAsnLys...AsnIleSerAsnTyrValLeuAspSerTyr 515
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1805 ACGATAACTCTAAAAGTGATACATCTCTAACTATTAAGCGTACGACTTAC 1854
516 GlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSe 532
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1855 GGAAAGTATGTACTTGTAGGTTTCCCTGTAATGAAATAAACTCCTATTA 1904
532 rPheGlyLeu...AsnAlaAspAsnThrLysLeuHisG 544
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1905 TGTAGGATTAGCTCATACCTATAATAAAATTAAGTAACTTTGCTCTAGAAT 1954
544 lyGlyArgPheMetGlyIleSerAsnValLysGlnLeuMetAlaAspGly 560
   : : : : : | : : : : | : : : : | : : : : | : : : : |
1955 ATAACCGTAATTTATATATTCAATCAATCAATGAAA... 1986
561 GlyLysIleGlnValAspAsnAsnGlyIleProAspPheLysHisAspTyr 577
   : : : : : | : : : : | : : : : | : : : : | : : : : |
1987 .....TTTAAAGGTAATGGCATTAATAAAACAAATGACTTTGATTT 2024
577 rThrThrTyrAsnAlaIleLeuGlyTyrAsnTyrSerSerLeuAspArgp 594
   : : : : : | : : : : | : : : : | : : : : | : : : : |
2025 TTCT.....TTTGGTGGAACTATAACAGCCTTAATAGAG 2059
594 roValPheProThrGlnGlyMetSerHisSerValasp...LeuThr 608
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2060 GCTATTTCCCAACTAAAGGGGTTAAAGCAAGCTTTGGTGGAGAGTTACT 2109
609 ValGlyPheGlyAspLysThrHisGlnLysValValTyrGlnGlyAsnIl 625
   : : : : : | : : : : | : : : : | : : : : | : : : : |
2110 ATTCAGGTTCTGTATAACAATACTACAAACTAAAGTCAGATGTACAGGG 2159
625 eTyrArgProPhe...IleLysLysSerValL 635
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2160 TTTCCTACCATATACAGAGATCACCTCTGGGTTGATCTCAGAAAGCAT 2209
635 euArgGlyTyrAlaLysLeuGlyTyrGlyAsn...AsnLeuProPheTyr 650
```



594 roValPheProThrGlnGlyMetSerHisServAlasp.....LeuThr 608  
:|||||::|||::|||::|||::|||::|||  
2060 GCATTTTCCCACTAAAGGGGTAAAGCAAGCTTTGGTGACGAAGTACT 2109  
:  
  
609 ValGlyPheGlyAspLysThrHisGlnLysValValTyrgInGlyAsnIle 625  
:::|||::|||::|||::|||::|||::|||::|||  
2110 ATTCAGAGTTCGATACAANAATTACAAAATAAGTCAGATGTACAGGG 2159  
:  
  
625 eTyArgPropHe.....IleLysLysServall 635  
:::|||::|||::|||::|||::|||::|||  
2160 TTCTACCACCATTAGACAGAGATCACCTCGTGGTTGTATCTGCAAAAGCAT 2209  
:  
  
635 euArgGlyTyralalysLeuGlyTyrgLyasn....AnLeuProPheTyrr 650  
:|||||::|||::|||::|||::|||::|||  
2210 CTCGAGGATATGCCAAT...GGTTTTGGAACAACAGCCTTTACCGTCTCAT 2256  
:  
  
651 GluAsnPhetyrAlacGlyTyrgLyservAlargGlyTyraspGlnSe 667  
:::|||::|||::|||::|||::|||::|||::|||  
2257 CAACCTTATACAGCGGTGCATCGGTCAATTACGTGTTTGCCTTAATGG 2306  
:  
  
667 rSerLeuglyProArgrserGlnAlatyrlThuThraalarGarglyGlnG 684  
:|||||::|||::|||::|||::|||::|||  
2307 TAGTATTGGACCTAACGCA.....ATTATCGCGAATATCGTAAATG 2347  
:  
  
684 InThrLeucUly.....GluvalValgLygly 693  
:::|||::|||::|||::|||::|||::|||::|||  
2348 GTAGTGGTACTGGTACTTTTAAAGAATAAGTCTGATGTATTGGTGGT 2397  
:  
  
694 AsnAlalaLeuAlathrpheGlysergluleuileLeuproPheLe 710  
:|||||::|||::|||::|||::|||::|||  
2398 AATGCAATCGCTACAGCTAGCGCAGAGTTAATTGTGCCAATCCTCATTTGT 2447  
:  
  
710 sClyAsprpile....AspgInvalarqprovalilpheheileclugly 726  
:|||||::|||::|||::|||::|||::|||::|||  
2448 GAGCGATAAGACCCAATAATACGGTCGCAACCTCTATTGTGTATCGCG 2497  
:  
  
726 lyGlnvalPheaSpthrThr...GlyMetaSplysGlnThrIleAspleu 741  
:|||||::|||::|||::|||::|||::|||  
2498 CAAAGTGTTTGGAATACTAAATGGAAAAATCAGATAAAATGGATTAGACAGC 2547  
:  
  
742 ThrGlnPheLysAspproGlnAlathralaGlughsnAlalysAlaal 758  
:|||||::|||::|||::|||::|||::|||  
2548 GATGTATTAAAA..... 2559  
:  
  
758 asNArgProLeuLeuthrGlnAsplsgLynLeuarTyrrSerAlaglv 775  
:|||||::|||::|||::|||::|||::|||  
2560 ..AGATTGCCTGATTATGCAAAATCAAGCCGATTTCGCGCCTCTACAGSTG 2608  
:  
  
775 alGlyAlathrTrpTyrrThrProileGlyProLeuSerIleSerrtyrAla 791  
:|||||::|||::|||::|||::|||::|||  
2609 TCGGATTCCAATGGCAAATCTCCTATTGGCCCATTTGGTATTCTCTATTATGCC 2658  
:  
  
792 LysProLeuAsnLysLysGlnAsnAspcGlnThrAspThrValGlnPheGI 808  
:|||||::|||::|||::|||::|||::|||  
2659 AAACCAATTAATAAATATGAANAATGATGATGTCGAACAGTTCCAATTAG 2708  
:  
  
808 nIleglyServalpHe 813  
:|||||::|||::|||::|||::|||::|||  
2709 TAATTGGAGGTCTTTC 2724

seq\_documentation\_block:  
LOCUS A38692 2984 bp  
DEFINITION Sequence 3 from Patent WO9412641.

KEYWORDS	SOURCE	ORGANISM	REFERENCE
	unidentified.		1 (bases 1 to 1000)
	unidentified		unclassified.
	unclassified.		



```
1651 TATTGGTTACGGTACAGAGAGTGGTATTAGTTATCAAGCAAGTGTAAAC 1700
452 InAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1701 AGATAATTCTTGGGAACAGAGCGGCAGTAGTATAGTGGTACGAAA 1750
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
469 SerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheTh 485
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1751 AATGATTATGGTACGAGTGTCAATTTGGGTATACCGAGCCCTATTTTAC 1800
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
485 rValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLysT 502
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1801 TAAAGATGGTGTGTA.....AGTCTTGGCGAAATGTTTCTTTGAAAACT 1844
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
502 YrAspAsnLys.....AsnIleSerAsnTyrValLeuAspSerTyr 515
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1845 ACGATAACTCTAAAGTGATACATCTCTAAGTATAAGCGTACGAGCTTAC 1894
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
516 GlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSe 532
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1895 GGAAGTAATGTACTTGTAGTTTCCCTGTAATGAAATAAATCACTCTATTA 1944
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
532 rPheGlyLeu.....AsnAlaAspAsnThrLysLeuHisG 544
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1945 TGTAGGATTAGGTACATACCTATAATAAATTAGTAACCTTTGCTCTAGAAT 1994
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
544 IyGlyArgPheMetGlyIleSerAsnValLysGlnLeuMetAlaAspGly 560
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1995 ATAACCGTAATTATATATCAATCAATGAAA..... 2026
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
561 GlyLysIleGlnValAspAsnAsnGlyIleProAspPheLysHisAspTy 577
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2027 .....TTTAAAGGTAATGGCATTAAACAAATGACTTTGATTT 2064
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
577 rThrThrTyrAsnAlaIleLeuGlyTrpAsnTyrSerSerLeuAspArgP 594
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2065 TTCT.....TTTGGTTGGAACTATAACAGCCTTAATAGAG 2099
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
594 roValPheProThrGlnGlyMetSerHisSerValAsp.....LeuThr 608
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2100 GCTATTTCCTCAACTAAAGGGTTAAAGCAAGTCTTGGTGGACGAGTTACT 2149
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
609 ValGlyPheGlyAspLysThrHisGlnLysValValTyrGlnGlyAsnI 625
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2150 ATTCCAGGTTCTGATAACAAATACTACAACCTAAAGTCAGATGTACAGG 2199
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
625 eTyrArgProPhe.....IleLysLysSerValL 635
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2200 TTCTACCATTTAGACAGATCACCTCTGGGTTGTATCTGCAAAAGCAT 2249
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
635 euArgGlyTyrAlaLysLeuGlyTyrGlyAsn...AsnLeuProPheTyr 650
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2250 CTGCAGGATATGCAAA...GGTTTGGAAACAAGCGTTTACCGTTCTAT 2296
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
651 GluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2297 CAAACTTATACAGCGGGTGGCATCGGTTCATTACGTGGTTTTTGTATGG 2346
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2347 TAGTATTGACCTAACGCA.....ATTATCCCGAATATGTAATG 2387
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
684 InThrThrLeuGly.....GluValValGlyGly 693
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2388 GTAGTGGTACTGGTACTTTTAAAGAGATAAGTCTGTATCTGATGTTGGT 2437
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
694 AsnAlaLeuAlaThrPheGlySerGlnLeuIleLeuProLeuProPheLy 710
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2438 AATGCAATCGCTACAGCTAGCGCAGAGTAATGTGCCAACTCCATTGT 2487
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
710 sGlyAspTrpIle...AspGlnValArgProValIlePheIleGluGlyG 726
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2488 GAGCGATAAGACCAAAATACGGTCCGAACCTCCTTATTGTGTATGCGG 2537
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```







OM of: US-09-701-711-2 to: GenEmbl.\* out\_format : pfs  
Date: Sep 19, 2002 4:41 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:  
-MODEL=framet.p2n.model -DEV=xlp  
-O=/cgn2.1/USPTO.spool/US09701711/runat\_17092002\_141431\_28888/app\_query.fasta\_1.882  
-DB=GenEmbl -QFMT=fastap -SUFFIX=sepl7oli.rge -GAPOP=4.500  
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000  
-FCAPOP=6.000 -FCAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=quality  
-THR\_MIN=1 -ALIGN=40 -MODE=LOCAL -OUTFMT=pfs -NORM=ext  
-HAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09701711@cgn1.1.8909 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:  
Query: US-09-701-711-2  
Query length: 813  
Database: GenEmbl.\*  
Database sequences: 1797656  
Database length: 1873333701  
Search time (sec): 3432.230000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000  
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score.list:	Strd Orig	ZScore	Escore	Len	Documentation
gb_pat:AX009287	+ 813.00	17143.58	0.0	2442	AX009287 Sequence 1 from Patent
gb_pat:AX009289	+ 749.00	15790.34	0.0	2442	AX009289 Sequence 3 from Patent
gb_pat:AX067462	+ 749.00	15762.52	0.0	99629	AX067462 Sequence 37 from Patent
gb_pat:AX081536	+ 36.00	1721.05	8.4e-32	1000	AX081536 Sequence 41 from Patent
gb_pat:AX009293	+ 11.00	213.42	0.0016	61	AX009293 Sequence 7 from Patent
gb_pat:AX009294	+ 11.00	212.72	0.0017	67	AX009294 Sequence 8 from Patent
gb_ba:PLU236920	+ 11.00	181.32	0.0976	4405	AX009294 Sequence 8 from Patent
gb_pat:AX078498	+ 10.00	166.77	0.6312	1830	AJ236920 Photorehabus luminesc
gb_ba:AF47013	+ 10.00	164.63	0.8301	2433	AX078498 Sequence 12 from Patent
gb_ba:AE005193	+ 10.00	163.85	0.9176	2700	AF47013 Escherichia coli O15
gb_ba:AE004297	+ 10.00	153.56	3.44	10649	AE005193 Escherichia coli O15
gb_ba:AE003941	- 10.00	153.38	3.51	10901	AE004297 Vibrio cholerae chrc
gb_ba:AE004784	- 10.00	152.84	3.76	11711	AE003941 Xylella fastidiosa S
gb_ba:AE000127	+ 10.00	152.37	4.00	12469	AE004784 Pseudomonas aerugin
gb_ba:AE008705	+ 10.00	151.05	4.74	14877	AE000127 Escherichia coli K12
gb_htg:AC104418	+ 10.00	147.78	7.20	22997	AE008705 Salmonella typhimur
gb_htg:AC1017607	+ 10.00	140.33	18.73	62088	AC104418 Homo sapiens chromos
gb_ba:ECOPSF	+ 10.00	138.31	24.26	81255	AC017607 Drosophila melanogas
gb_ba:ECUT0214	+ 10.00	137.43	27.18	91430	D83536 Escherichia coli genom
gb_in:AC009366	+ 10.00	135.19	36.20	123171	U70214 Escherichia coli chrc
gb_ba:AJ414154	+ 10.00	132.78	49.31	169839	AC009366 Drosophila melanoga
gb_ba:AJ414154	+ 10.00	132.36	52.04	179611	AC023676 Drosophila melanoga
gb_ba:AJ414146	+ 10.00	131.41	58.84	204051	AJ414154 Versinia pestis str
gb_ba:AL627266	+ 10.00	131.19	60.50	210050	AJ414146 Versinia pestis str
gb_ba:AP002550	+ 10.00	129.36	76.49	268050	AL627266 Salmonella enterica
gb_pat:AX003597	+ 10.00	128.99	80.19	281530	AP002550 Escherichia coli O1
gb_pat:AX078597	+ 10.00	128.63	84.03	295566	AX003597 Drosophila melanoga
gb_pr:AY062939	+ 9.00	144.88	10.45	2019	AX078597 Sequence 111 from Pat
gb_pl:2MZAG2	- 9.00	143.53	12.43	2418	AY062939 Homo sapiens REFSDR2
gb_pl:SPCC188	- 9.00	135.48	34.88	7071	X80206 Z.mays zag2 gene. 9/199
gb_pl:SPAC84	+ 9.00	123.40	164.33	35412	AL049662 S.pombe chromosome I
gb_pl:SPCC584	+ 9.00	122.84	176.49	38141	Z66569 S.pombe chromosome I
gb_htg:AC102899	- 9.00	122.15	192.76	41803	AL032824 S.pombe chromosome I
gb_htg:AC102899	- 9.00	118.92	291.87	64341	AC102899 Mus musculus clone R
gb_pr:AL359454	- 9.00	116.00	424.39	94948	AC095320 Rattus norvegicus cl
gb_htg:AC094530	- 9.00	115.48	453.94	10830	AL359454 Human DNA sequence
gb_htg:AC094530	- 9.00	115.11	476.08	106997	AC094530 Rattus norvegicus c
gb_htg:AC087596	+ 9.00	114.95	485.55	109210	AC087596 Oryza sativa chromid

gb\_htg:LMFLCHR16\_09 - 9.00 114.90 488.93 110000 ! Continuation (10 of 11) o  
gb\_htg:AC016033 - 9.00 114.39 521.78 117693 ! AC016033 Homo sapiens chr  
gb\_pr:AL512292 + 9.00 113.34 597.05 135889 ! AL512292 Human DNA sequen  
gb\_pl:AP003451 - 9.00 112.57 658.99 150019 ! AP003451 Oryza sativa chr  
gb\_htg:AL161895 + 9.00 112.39 673.97 153566 ! AL161895 Homo sapiens chr  
gb\_htg:AC106473 + 9.00 112.27 684.96 156169 ! AC106473 Rattus norvegicu  
gb\_htg:AC018550 - 9.00 111.96 712.95 162807 ! AC018550 Homo sapiens chr  
  
seq\_name: gb\_pat:AX009287  
seq\_documentation\_block:  
LOCUS AX009287 2442 bp DNA linear PAT 06-SEP-2000  
DEFINITION Sequence 1 from Patent WO9963093.  
ACCESSION AX009287  
VERSION AX009287.1 GI:9996618  
KEYWORDS  
SOURCE Moraxella catarrhalis.  
ORGANISM Moraxella catarrhalis  
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;  
Moraxella.  
  
REFERENCE 1 (bases 1 to 2442)  
AUTHORS Vinals-Bassols,C.  
TITLE Bas027 proteins and genes from moraxella catarrhalis, antigens,  
antibodies, and us  
JOURNAL Patent: WO 9963093-A 1 09-DEC-1999;  
SMITHKLINE BEECHAM BIOLOG (BE); VINALS BASSOLS CARLOTA (BE)  
FEATURES  
source Location/Qualifiers  
i. 2442  
/organism="Moraxella catarrhalis"  
/db\_xref="taxon:480"  
BASE COUNT 737 a 461 c 533 g 711 t  
ORIGIN

alignment_scores:	Quality:	813.00	Length:	813
	Ratio:	1.000	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000	
alignment_block:	US-09-701-711-2	x AX009287		
Align seg 1/1 to: AX009287	from: 1	to: 2442		
1 MetArgAsnSerTyrPheLysGlyPheGlnValSerAlaMetThrMetAl 17				
1 ATGCGTAATTCATATTTAAAGGTTTTCAGGTGAGTGCATGCAATGCAATGCC 50				
17 aValMetMetValMetSerThrHisAlaGlnAlaAlaAspPheMetAla 34				
51 TGTCTATGATGTAATGTCACTCATGTCACAGCGCGGATTTTATGGCAA 100				
34 sAspPheThrIleThrGlyLeuGlnArgValThrIleGlnSerLeuGln 50				
101 ATGACATTTACCATCACAGGACTACAGCGAGTGACCATTTCAAGAGCTTACAA 150				
51 SerValLeuProPheArgLeuGlyGlnValValSerGluAsnGlnLeuAl 67				
151 AGCGTGTCGCGTTCGTTGGTCAAGTGGTGACGGAACCAAGCTGGC 200				
67 aAspGlyValLysAlaLeuTyrAlaThrGlyAsnPheSerAspValGlnV 84				
201 TGATGGTGTCAAAGCACATTTATGCAACAGGCAATTTTTCAGATGTGCAAG 250				
84 alTyrHisGlnGluGlyArgIleLeuTyrGlnValThrGluArgProLeu 100				
251 TCTATCATCAAGAGGCGGTATCATCTATCAGTAACCGAAAGGCGGTGA 300				
101 IleAlaGluIleAsnPheGluGlyAsnArgLeuIleProLysGluGlyLe 117				
301 ATCGCTGAGATTAATTTTGGGGCAATCGCTTAATTCAAAAGAGGTTCT 350				
117 uGlnGluGlyLeuLysAsnAlaGlyLeuAlaValGlyGlnProLeuLysG 134				

|||||  
351 ACAAGAGGGCTAAAAAATGCTGGCTTAGCTGGGTCAACCACCTAAAAAC 400  
|||||  
134 lAlaThrValGlnMetIleGluThrGluLeuThrAsnGlnThrIleSer 150  
|||||  
401 AGCCACAGTACAGATGATCGAAACGAGCTTACCAATCAATATATATCA 450  
|||||  
151 GlnGlyTyrTyrAsnThrGluIleThrValLysGlnThrMetLeuAspG1 167  
|||||  
451 CRAAGGCTATTATAATACCGAATTAATCTGTCAACACAGCAGTCTTGATGG 500  
|||||  
167 yAsnArgValLysLeuAspMetThrPheLaGluGlyLysProAlaArgV 184  
|||||  
501 TAACTCGTGTAAAGCTTGATATGACCTTGTCTGAAGGTAAACCTGACGGG 550  
|||||  
184 aValAspIleAsnIleIleGlyAsnGlnHisPheSerAspAlaAspLeu 200  
|||||  
551 TGGTTGATATTATATCATTTGGCAATCAGCATTTTAGCGATCGAGATTG 600  
|||||  
201 IleAspValLeuAlaIleLysAspAsnLysIleAsnProLeuSerLysAl 217  
|||||  
601 ATTGATGCTGCTTGGGATTAAGGATATAAATCAATCCACTGCTAAAGC 650  
|||||  
217 aAspArgTyrThrGlnGluLysLeuValThrSerLeuGluAsnLeuArgA 234  
|||||  
651 TGAACCGTTATACTCAAGAAAGCTGTGCACCACTTTAGAGAAATTTGCGTG 700  
|||||  
234 lAlaTyrLeuAsnAlaGlyPheValArgPheGluIleLysAspAlaLys 250  
|||||  
701 CTAAATATCTCAATGACGGGTTTGTGCGTTTGGAGATTAAAGATGCTAAG 750  
|||||  
251 LeuAsnIleAsnGluAspLysAsnArgIlePheValGluIleSerLeuH1 267  
|||||  
751 CTAAATATTAAATGAAGATAAAACCGTATCTTTGTTGAGATTTCATTGCA 800  
|||||  
267 sGluGlyGluGlnTyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeut 284  
|||||  
801 TGAAGGTGAGCAATATCGCTTTGGACAGACACAGATTTTGGGTAATTTAA 850  
|||||  
284 hTyrThrGlnAlaGluLeuGluAlaLeuLeuLysPheLysAlaGluGlu 300  
|||||  
851 CTTATATCTCAAGCAGAACTTGAGGCACTGCTTAAATTCAAACGAGAA 900  
|||||  
301 GlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLy 317  
|||||  
901 GGGTTTTCACAGCCATGCTTGAGCAAAACAACAATATCAGTACCAA 950  
|||||  
317 sPheGlyAspAspGlyTyrTyrThrAlaGlnIleArgProValThrArgI 334  
|||||  
951 ATTTGTGACGATGGCTATTATTATGCTCAAAATCCGTCCTGTAAACACGCA 1000  
|||||  
334 leAsnAspGluSerArgThrValAspValGluTyrTyrIleAspProVal 350  
|||||  
1001 TTAATCATGAAAGCTGCTACGGTTGATGGGAATATTATATGACCCTGTA 1050  
|||||  
351 HisProValTyrValArgArgIleAsnPheThrGlyAsnPhelystThrG1 367  
|||||  
1051 CACCCCTGCTATGACCGCGTATTAATTTTACAGGTAACCTTTAAGACCCA 1100  
|||||  
367 nAspGluValLeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlas 384  
|||||  
1101 AGATGAAGTACTCCGTCGTGAGATGCGCAACTTGAAGGTGCGCTTGGCAT 1150  
|||||  
384 erAsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhe 400  
|||||  
1151 CTAATCAAAAAATCCAGCTGCTCGTCGACGCTTGATGCGGACTGGGTTT 1200  
|||||  
401 PheLysHisValThrValAspThrArgProValProAsnSerProaspG1 417  
|||||  
1201 TTTAAACATGTTTACCGTTGATCTCGCAGTACCCCACTCAGCTGATCA 1250  
|||||  
417 nValAspValAsnPheValValGluGlnProSerGlySerSerThrI 434  
|||||

1251 GGTGTGATGTAATTTTGTGGTTGAAGAACACCTTCAGGATCATCAACCA 1300  
|||||  
434 leAlaAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspVal 450  
|||||  
1301 TCGCAGCAGGCTACTCTCAAGTGGTGGTGAACCTTTTCAATTTGATGTT 1350  
|||||  
451 SerGlnAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSe 467  
|||||  
1351 TCTCAAAATAACTTTTATGGGTACAGTAAGCAGCTCAATGCTTCGTTTTC 1400  
|||||  
467 rArgSerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrP 484  
|||||  
1401 TCGCTCTGAGACCCGCTGAGGTGTATAGTTTGGGTATGACCAACCCATACT 1450  
|||||  
484 heThrValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThr 500  
|||||  
1451 TTACCGTAAATGGCGTCTCGAAAGCTTGAGTGGCTACTATCGTAAACC 1500  
|||||  
501 LysTyrAspAsnLysAsnIleSerAsnTyrValLeuAspSerTyrGlyG1 517  
|||||  
1501 AAGTATGATAACAAGAACAATTAGTAATTTATGTACTTCTTATGTTGGTGG 1550  
|||||  
517 ySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSerPheG 534  
|||||  
1551 CTCATTTAAGCTATGGATATCCAAATTCATGAAATCAACGCATTAAGCTTTG 1600  
|||||  
534 lYLeuAsnAlaAspAsnThrLysLeuHisGlyGlyArgPheMetGlyIle 550  
|||||  
1601 GTCGTGATGCTGACAAATACCAAGCTTCATGGCGTCTGTTTATGCGCAT 1650  
|||||  
551 SerAsnValLysGlnLeuMetAlaAspGlyGlyLysIleGlnValAspas 567  
|||||  
1651 AGTAATGTCAAGCAGCTGATGGCAGATGGTGCAAAATTCAGTGGATAA 1700  
|||||  
567 nAsnGlyIleProAspPheLysHisAspTyrThrThrTyrAsnAlaIle 584  
|||||  
1701 TAAATGCTTCTGATTTTAAAGCATGATTACACACCTTACATGCCATTT 1750  
|||||  
584 euGlyTrpAsnTyrSerSerLeuAspArgProValPheProThrGlnGly 600  
|||||  
1751 TGGGTGGAATTTATCAAGTCTAGATCGCCCTGTATTTCACCAACCAAGGC 1800  
|||||  
601 MetSerHisSerValAspLeuthrValGlyPheGlyAspLysThrHisG1 617  
|||||  
1801 ATGAGTCACTCTGTAGATTTGACGGTTGGTGGTGGTGGTGGTGGTGGTGGT 1850  
|||||  
617 nLysValValTyrGlnGlyAsnIleTyrArgProPheIleLysLysServ 634  
|||||  
1851 AAAAGTGGTTTATCAAGGCAATATCTATCGCCCATTTTATCAAAAAATCAG 1900  
|||||  
634 aLeuArgGlyTyrAlaLysLeuGlyTyrGlyAsnAsnLeuProPheTyr 650  
|||||  
1901 TCTTGGTGGATACGCCAAAGTTAGGCTATGGCAATAATTTTACCATTTTAT 1950  
|||||  
651 GluAsnPheTyrAlaGlyTyrGlySerValArgGlyTyrAspGlnSe 667  
|||||  
1951 GAAATTTCTATGACGGGCTATGTTTGGTTCGGTGGTGGTGGTGGTGGTGGT 2000  
|||||  
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684  
|||||  
2001 CTCCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2050  
|||||  
684 lnrThrThrLeuGlyGluValValGlyGlyAsnAlaLeuAlaThrPheGly 700  
|||||  
2051 AAACCACTAGGAGAGGTTGTTGGTGGTAAATGCTTTGGCAACTTTTCGGC 2100  
|||||  
701 SerGluLeuIleLeuProLeuProPheLysGlyAspTrpIleAspGlnVa 717  
|||||  
2101 AGTGAGCTGATTTTACCTTTGCCATTTAAAGGTGATTGGATGATCAGGT 2150  
|||||  
717 lArgProValIlePheIleGluGlyGlnValPheAspThrThrGlyM 734  
|||||  
2151 GCGTCCAGTGATATTTCATTCAGGCGGCTCAGGTTTGTGATACACAGGTA 2200  
|||||

734 etAspLysGlnThrIleAspLeuThrGlnPheLysAspProGlnAlaThr 750  
 |||||  
 2201 TGGATAACAACACCATGATTAAACCAATTTAAAGACCCACACACA 2250  
 751 AlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuThrGlnAspLys 767  
 |||||  
 2251 GCTGAACAAAATGCAAAAGCAGCCCAATCGCCGCTACTAACCCAAAGATA 2300  
 767 sgInLeuArgTyrSerAlaGlyValGlyAlaThrTrpTyrThrProIleG 784  
 |||||  
 2301 ACAGTTGGCTTATAGTGTGGTGGTGGCAACTTGGTATAGCCCAATG 2350  
 784 lyProLeuSerIleSerTyrAlaLysProLeuAsnLysLysGlnAsnAsp 800  
 |||||  
 2351 GTCCCTTATCTATTAGCTATGCCAAGCCATTGAATAAAAAACAATAATGAT 2400  
 801 GlnThrAspThrValGlnPheGlnIleGlySerValPhe 813  
 |||||  
 2401 CAGACCGATACGGTACAGTTCAGATTGGTAGTGTCTTT 2439

seq\_name: gb\_pat:AX009289

seq\_documentation\_block:  
 LOCUS AX009289 2442 bp DNA linear PAT 06-SEP-2000  
 DEFINITION Sequence 3 from Patent WO9963093.  
 ACCESSION AX009289  
 VERSION AX009289.1 GI:9996619

KEYWORDS Moraxella catarrhalis.  
 SOURCE Moraxella catarrhalis  
 ORGANISM Moraxella catarrhalis  
 Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;  
 Moraxella.

REFERENCE 1 (bases 1 to 2442)  
 AUTHORS Vinals-Bassols,C.  
 TITLE Bas027 proteins and genes from moraxella catarrhalis, antigens,  
 antibodies, and uses

JOURNAL Patent: WO 9963093-A 3 09-DEC-1999;  
 SMITHKLINE BEECHAM BIOLOG (BE); VINALS BASSOLS CARLOTA (BE)

FEATURES  
 Location/Qualifiers  
 source  
 1..2442  
 /organism="Moraxella catarrhalis"  
 /db\_xref="taxon:480"

BASE COUNT 735 a 461 c 535 g 711 t  
 ORIGIN

alignment\_scores:  
 Quality: 749.00 Length: 749  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-701-711-2 x AX009289 ..

Align seg 1/1 to: AX009289 from: 1 to: 2442

65 GlnLeuAlaAspGlyValLysAlaLeuTyrAlaThrGlyAsnPheSerAs 81  
 |||||  
 193 CAGTTGGCTGATGGTGTCAAGACACTTTATGCAACAGGCAATTTTCAGA 242  
 81 pValGlnValTyrHisGlnGluGlyArgIleIleTyrGlnValThrGlu 98  
 |||||  
 243 TGTGCACTCTCATCATCAAGAGGCGGTATCATCTATCAGGTAAACCGAAA 292  
 98 rgProLeuIleAlaGluIleAsnPheGluGlyAsnArgLeuIleProLys 114  
 |||||  
 293 GGCCGTTAATCGCTGAGATTAATTTGAGGCAATCGCTTAATCCAAAA 342  
 115 GluGlyLeuGlnGluGlyLeuLysAsnAlaGlyLeuAlaValGlyGlnPr 131  
 |||||  
 343 GAAGGCTACAAAGAGGGCTAAAAAATGCTGGCTAGCTGTGGGTCAACC 392

131 oLeuLysGlnAlaThrValGlnMetIleGluThrGluLeuThrAsnGlnT 148  
 |||||  
 393 ACTAAACAAGCCACAGTACAGATGATCGAAACCGAGCTTACCATCAAT 442  
 148 yrIleSerGlnGlyTyrTyrAsnThrGluIleThrValLysGlnThrMet 164  
 |||||  
 443 ATATATCAACAAGGCTATTATATACCGAAATTTACTGTCAACACAGCATG 492  
 165 LeuAspGlyAsnArgValLysLeuAspMetThrPheAlaGluGlyLysPr 181  
 |||||  
 493 CTTGATGGTAATCGTGTAAAGCTTGATATGACCTTTGCTGAAGGTAAACC 542  
 543 TGCACGGGTGGTGTGATATTATATCATTTGGCAATCAGCATTTTAGCGATG 592  
 198 laAspLeuIleAspValLeuAlaIleLysAspAsnLysIleAsnProLeu 214  
 |||||  
 593 CAGATTTGATTGATGTCTTGGATTAAAGGATAATAAATCAATCAACTG 642  
 215 SerLysAlaAspArgTyrThrGlnLysLeuValThrSerLeuGluAs 231  
 |||||  
 643 TCTAAAGCTGACCGTTATCTCAACAAAAGCTGTGACCACTTTAGAGAA 592  
 231 nLeuArgAlaLysTyrLeuAsnAlaGlyPheValArgPheGluIleLysA 248  
 |||||  
 693 TTTGCGTGTCTAAATATCTCAATGCAGGGTTTGTGCGTTTGGATTAAG 742  
 248 spAlaLysLeuAsnIleAsnGluAspLysAsnArgIlePheValGluIle 264  
 |||||  
 743 ATGCTAAAGCTTAATATTAAAGATAAAACCGTATCTTTGTTGAGATT 792  
 265 SerLeuHisGluGlyGluGlnTyrArgPheGlyGlnThrGlnPheLeuGl 281  
 |||||  
 793 TCATTGCAATGAAGGTGAGCAATATCGCTTTGGACAGACACACTTTTGG 842  
 281 yAsnLeuThrTyrThrGlnAlaGluLeuGluAlaLeuLysPheLysA 298  
 |||||  
 843 TAATTTAACTTATCTCAAGCAGAACTTTGAGGCACTGCTTAAATTCAAA 892  
 298 laGluGluGlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsnIle 314  
 |||||  
 893 CAGAAAGAGGGTTTTCACAAAGCCATGCTTGAGCAACAACAACAATATC 942  
 315 SerThrLysPheGlyAspAspGlyTyrTyrAlaGlnIleArgProVa 331  
 |||||  
 943 AGTACCAAAATTTGGTACGATGGCTATTATATGCTCAAAATCCGCTCTGT 992  
 331 lThrArgIleAsnAspGluSerArgThrValAspValGluTyrTyrIleA 348  
 |||||  
 993 AACAGCATTAATGATGAAAGTCGTACGGTTGATGTGGAATATTATATTG 1042  
 348 spProValHisProValTyrValArgArgIleAsnPheThrGlyAsnPhe 364  
 |||||  
 1043 ACCCTGTACACCCCTGCTCTATGTCGCCGCTATTAAATTTTACAGSTAAC 1092  
 365 LysThrGlnAspGluValLeuArgArgGluMetArgGlnLeuGluGlyAl 381  
 |||||  
 1093 AAGACCCCAAGATGAAGTACTCCGTCGTGAGATGCCAACAACCTTGAAG 1142  
 381 aLeuAlaSerAsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArgT 398  
 |||||  
 1143 GTTGCATCTAATCAAAAAATCCACGCTGCTCGTCACGCTTGATGCGGA 1192  
 398 hrGlyPhePheLysHisValThrValAspThrArgProValProAsnSer 414  
 |||||  
 1193 CTGGGTTTTTAAACATGTTACCGTTGATCTCGTCCAGTACCCCACTCA 1242  
 415 ProAspGlnValAspValAsnPheValValGluGlnProSerClySe 431  
 |||||  
 1243 CCTGATCAGGTTGATGTAATTTTGTGGTTGAAGAACCAACCTTCAGGATC 1292  
 431 rSerThrIleAlaAlaGlyTyrSerGlnSerGlyValThrPheGlnP 448

```
|||||
1293 ATCAACCATCGCAGCAGGTACTCTCAAAAGTGGTGTAACTTTTCAAT 1342
|||||
448 heAspValSerGlnAsnAsnPhMetGlyThrGlyLysHisValAsnAla 464
|||||
1343 TTGATGTTTCTCAAAATAACTTTATGGGTACAGTAAGCAGCGTCAATGCT 1392
|||||
465 SerPheSerArgSerGluThrArgGluValTyrSerLeuGlyMetThrAs 481
|||||
1393 TCGTTTCTCGCTCTGAGACCCGTGAGGTGTATAGTTTGGGTATGACCAA 1442
|||||
481 nProTyrPheThrValAsnGlyValSerGlnSerLeuSerGlyTyrTyra 498
|||||
1443 CCCATACTTTACCGTAAATGGCGCTCTCGCAAGCTTGAGTGGCTACTATC 1492
|||||
498 rgLysThrLysTyrAspAsnLysAsnLysSerAsnTyrValLeuAspSer 514
|||||
1493 GTAAACCAAGTAGTATGAACAAGAACATAGTAATTATGCTACTGTATCT 1542
|||||
515 TyrGlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIl 531
|||||
1543 TATGGTGGCTCATAGCTATGATATCCAAATGATGAATAATCAACGCAT 1592
|||||
531 eSerPheGlyLeuAsnAlaAspAsnThrLysLeuHisGlyGlyArgPheM 548
|||||
1593 AAGCTTTGGTCTCAATGCTGACAATACCAAGCTTCATGGCGGTCTTTTA 1642
|||||
548 etGlyIleSerAsnValLysGlnLeuMetAlaAspGlyGlyLysIleGln 564
|||||
1643 TGGGCATTAGTAATGTCACAGCAGCTGATGGCAGATGGTGCAAAATTCAA 1692
|||||
565 ValAspAsnAsnGlyIleProAspPheLysHisAspTyrThrTyrAs 581
|||||
1693 GTGGATAAATAATGCAATCTCTGATTTTAAGCATGATTACACAACCTACAA 1742
|||||
581 nAlaIleLeuGlyTyrPasnTyrSerSerLeuAspArgProValPheProT 598
|||||
1743 TGGCATTTTGGGTGGAAATATTCAAGTCTAGATCGCCCTGTATTTCCAA 1792
|||||
598 hrGlnGlyMetSerHisSerValAspLeuThrValGlyPheGlyAspLys 614
|||||
1793 CCCAAGCATGAGTCATCTGTAAGATTTGACGGTTGGTTTGGTGATAAA 1842
|||||
615 ThrHisGlnLysValValTyrGlnGlyAsnIleTyrArgProPheIleLy 631
|||||
1843 ACTCATCAAAAGTGGTTTATCAAGCAATATCTATCGCCATTTATCAA 1892
|||||
631 sLysSerValLeuArgGlyTyrAlaLysLeuGlyTyrGlyAsnAsnLeup 648
|||||
1893 AAATCAGTCTTGGGTGATACGCCAAGTTAGGCTATGGCAATAATTATC 1942
|||||
648 roPheTyrGluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyr 664
|||||
1943 CATTTATGAAATTTCTATGACGGCGGTATGGTTCGGTTCGGCTAT 1992
|||||
665 AspGlnSerSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgAr 681
|||||
1993 GATCAATCCCTCTTTGGGTCCACGCTCACAAAGCCTATTGACAGCTCGTCG 2042
|||||
681 qGlyGlnGlnThrThrLeuGlyGluValValGlyGlyAsnAlaLeuAlat 698
|||||
2043 TGGTCAACAAACCACTAGGAGAGGTGTGGTGGTAATGCTTTGGCAA 2092
|||||
698 hrPheGlySerGluLeuIleLeuProLeuProPheLysGlyAspTrpIle 714
|||||
2093 CTTTCGGCAGTAGCTGATTTTACCTTTGGCATTTAAAGGTGATTGGATA 2142
|||||
715 AspGlnValArgProValIlePheIleGluGlyGlnValPheAspTh 731
|||||
2143 GATCAGGTGGCTCAGTGTATTCATTGAGGCGGTGAGGTTTGTGATAC 2192
|||||
731 rThrGlyMetAspLysGlnThrIleAspLeuThrGlnPheLysAspProG 748
|||||
```

```
2193 AACAGGTATGGATAAAACAAACCATTTGATTTTAAACCAATTTAAAGACCCAC 2242
|||||
748 InAlaThrAlaGluGlnAsnAlaLysAlaAsnArgProLeuLeuThr 764
|||||
2243 AAGCAACAGCTGAACAAATGCAAAAGCAGCCAAATCGCCGCTACTAACCC 2292
|||||
765 GlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrTrpTyrTh 781
|||||
2293 CAAGATAAACAGTTGCTTATAGTGTGGTGGTGGTCAACTTGGTATAC 2342
|||||
781 rProLleGlyProLeuSerIleSerTyrAlaLysProLeuAsnLysLysG 798
|||||
2343 GCCCATTTGGTCTTTATCTATTAGCTATGCAAGCCATTGAATAAAAAAC 2392
|||||
798 InAsnAspGlnThrAspThrValGlnPheGlnIleGlySerValPhe 813
|||||
2393 AAAATGATCAGACCGATACGGTACAGTTCCAGATTGGTAGTGTCTTT 2439
|||||
seq_name: gb_pat:AX067462
seq_documentation_block:
LOCUS AX067462 99629 bp DNA linear PAT 24-JAN-2001
DEFINITION Sequence 37 from Patent WO0078968.
ACCESSION AX067462
VERSION AX067462.1 GI:12545082
KEYWORDS
SOURCE Moraxella catarrhalis.
ORGANISM Moraxella catarrhalis
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
Moraxella.
REFERENCE
1 (bases 1 to 99629)
AUTHORS Lagace,R.E., Patterson,C. and Berg,K.L.
TITLE Nucleotide sequences of moraxella catarrhalis genome
JOURNAL Patent: WO 0078968-A 37 28-DEC-2000;
Incyte Genomics, Inc. (US)
FEATURES
location/Qualifiers
source 1..99629
/organism="Moraxella catarrhalis"
/db_xref="taxon:480"
78467
unseq /note="a or g or c or t, unknown, or other"
BASE COUNT 29233 a 19222 c 21909 g 29264 t 1 others
ORIGIN
alignment_scores:
Quality: 749.00 Length: 749
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-701-711-2 x AX067462 ..
Align seg 1/1 to: AX067462 from: 1 to: 99629
65 GlnLeuAlaAspGlyValLysAlaLeuTyrAlaThrGlyAsnPheSerAs 81
|||||
66325 CAGTTGGCTGATGGTGTCAAAAGCACTTTATGCAACAGGCAATTTTCAGA 66374
|||||
81 pValGlnValTyrHisGlnGluGlyArgIleIleTyrGlnValThrGluA 98
|||||
66375 TGTGCAAGTCTATCATCAAGAGGGCGTATCATCTATCAGGTACCCGAA 66424
|||||
98 rgProLeuIleAlaGluIleAsnPheGluGlyAsnArgLeuIleProLys 114
|||||
66425 GGCGGTAAATCGCTGAGATTAATTTGAGGCAATCGCTTAATTCACAAA 66474
|||||
115 GluGlyLeuGlnGlyLeuLysAsnAlaGlyLeuAlaValGlyGlnPr 131
|||||
66475 GAAGTCTACAAGAAGGGCTTAAAAAATGCTGGCTTAGCTTGGGTCAACC 66524
|||||
131 oLeuLysGlnAlaThrValGlnMetIleGluThrGluLeuThrAsnGlnT 148
|||||
```

66525 ACTAAACAAAGCCACAGTACAGATGATCGAAACCGAGCTTACCAATCAAT 66574  
 148 yrIleSerGlnGlyTyrThrAsnThrGluIleThrValLysGlnThrMet 164  
 66575 ATATATCACAGGCTATTATATACCGAATTTACTGTCAACACACAGGATG 66624  
 165 LeuAspGlyAsnArgValLysLeuAspMetThrPheAlaGluGlyLysPr 181  
 66625 CTTGATGCTAATCGTGTAAAGCTTGATATGACCTTTGCTGAAGGTAAACC 66674  
 181 oAlaArgValValAspIleAsnIleLeGlyAsnGlnHisPheSerAspA 198  
 66675 TGCACGGGTGGTTGATATTAATATCATTTGGCAATCAGCATTTTACGATG 66724  
 198 laAspLeuIleAspValLeuAlaIleLysAspAsnLysIleAsnProLeu 214  
 66725 CAGATTTGATTTGATGCTGCTGCGATTAAAGGATAATAAATCAATCCACTG 66774  
 215 SerLysAlaAspArgTyrThrGlnGluLysLeuValThrSerLeuGluAs 231  
 66775 TCTAAAGCTGACCTTATCTCAAGAAAGCTGGTGACCACTTTAGAGAA 66824  
 231 nLeuArgAlaLysTyrLeuAsnAlaGlyPheValArgPheGluIleLysA 248  
 66825 TTTGCGTGCTAAATATCTCAATGACGGTTTGTGCGTTTGTGAGATTAAAG 66874  
 248 spAlaIleAsnIleAsnGluAspLysAsnArgIlePheValGluIle 264  
 66875 ATGCTAGCTTAAATATTAATGAAGATAAAACCGTATCTTTGTGAGATT 66924  
 265 SerLeuHisGluGlyGluGlnTyrArgPheGlyGlnThrGlnPheLeuGl 281  
 66925 TCATTGTCATGAAGTGAGCAATATCGCTTTGGACAGACACAGTTTGGG 66974  
 281 yAsnLeuThrTyrThrGlnAlaGluLeuGluAlaLeuLeuLysPheLysA 298  
 66975 TAATTTAACTATATCTCAAGCAGAACTTGAGGCACCTGCTTAAATTTCAAAG 67024  
 298 laGluGluGlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsnIle 314  
 67025 CAGAAGAAGGTTTTCACAGCCATGCTTGAGCAAAACAAACAATATC 67074  
 315 SerThrLysPheGlyAspAspGlyTyrTyrAlaGlnIleArgProVa 331  
 67075 AGTACCAAAATTTGTCGACGATGCTATTATATGCTCAAAATCCGCTCGT 67124  
 331 lThrArgIleAsnAspGluSerArgThrValAspValGluTyrTyrIleA 348  
 67125 AACACGCAATTAATGATGAAAGTCGTACGGTTGATGTGGAATATATATTG 67174  
 348 spProValHisProValTyrValArgArgIleAsnPheThrGlyAsnPhe 364  
 67175 ACCCTGTACACCCCTGTCTATGTACCGCATTAATTTTACAGGTAACTTT 67224  
 365 yLysThrGlnAspGluValLeuArgArgGluMetArgGlnLeuGluGlyAl 381  
 67225 AAGACCAAGATGAAGTACTCCGTCGTGAGATGCGACAACCTTGAAGTGC 67274  
 381 aLeuAlaSerAsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArgT 398  
 67275 GTTGGCATCTAAATCAAAAAATCCAGCTGTCTCGTGCACGCTTGATGCGGA 67324  
 398 hrGlyPhePheLysHisValThrValAspThrArgProValProAsnSer 414  
 67325 CTGGGTTTTTAAACATGTTTACCGTTGATCTCGTCCAGTACCCCACTCA 67374  
 415 ProAspGlnValAspValAsnPheValValIcIuGlnGlnProSerGlySe 431  
 67375 CCTGATCAGGTTGATGTAAATTTTGTGGTTGAGAAACAACACCTTCAGGATC 67424  
 431 rSerThrIleAlaAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnP 448  
 67425 ATCAACCATCGCAGCGGCTACTCTCAAGTGGTGGTGTAACTTTTCAAT 67474

448 heAspValSerGlnAsnAsnPheMetGlyThrGlyLysHisValAsnAla 464  
 67475 TTGATGTTTTCAAAATAACTTTATGGGTACAGTAACGACGCTCAATGCT 67524  
 465 SerPheSerArgSerGluThrArgGluValTyrSerLeuGlyMetThrAs 481  
 67525 TCGTTTTCTCGCTCTGAGACCCGTCGAGGTATAGTTTGGGTATGACCA 67574  
 481 nProTyrPheThrValAsnGlyValSerGlnSerLeuSerGlyTyrTyrA 498  
 67575 CCCATACCTTTACCGTAAATGGGCTCTCGCAAGCTTGAGTGGCTACTATC 67624  
 498 rGlyLysThrLysTyrAspAsnLysAsnIleSerAsnTyrValLeuAspSer 514  
 67625 GTAAACCAAGTATGATAACAAGAACAATTAGTAATTATGACTTGATTCT 67674  
 515 TyrGlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIl 531  
 67675 TATGTTGGCTCATTAAGCTATGGATATCCAATTGATGAAAAATCAACGCAT 67724  
 531 eSerPheGlyLeuAsnAlaAspAsnThrLysLeuHisGlyGlyArgPheM 548  
 67725 AAGCTTTGGTCTGAATGCTGACAATACCAAGCTTCATGGCGGTCGTTTTA 67774  
 548 etGlyIleSerAsnValLysGlnLeuMetAlaAspGlyGlyLysIleGln 564  
 67775 TGGCATTAGTAATGTCAAGCAGCTGATGGCAGATGGTGGCAAAATTCAA 67824  
 565 ValAspAsnAsnGlyIleProAspPheLysHisAspTyrThrTyrAs 581  
 67825 GTGGATAATAATGTCATTTCCTGATTTTAAAGCATGATTACACAACCTACA 67874  
 581 nAlaIleLeuGlyTyrAsnTyrSerSerLeuAspArgProValPheProT 598  
 67875 TGCCATTTGGGGTGAATATTCAAGTCTAGATCGCCCTGTATTTCCAA 67924  
 598 hrGlnGlyMetSerHisSerValAspLeuThrValGlyPheGlyAspLys 614  
 67925 CCCAAGGCATGAGTCATCTCTAGATTGACGGTTGGTTTGGTGATAAA 67974  
 615 ThrHisGlnLysValValTyrGlnGlyAsnIleTyrArgProPheIleLy 631  
 67975 ACTCATCAAAAAGTGGTTTATCAAGGCAATATCTATCGCCCATTTATCAA 68024  
 631 sLysSerValLeuArgGlyTyrAlaLysLeuGlyTyrGlyAsnAsnLeuP 648  
 68025 AAAATCAGTCTTGGCTGGATACGCCAAGTTAGGCTATGGCAATAATTTAC 68074  
 648 roPheTyrGluAsnPheTyrAlaGlyTyrGlySerValArgGlyTyr 664  
 68075 CATTTTATGAAAAATTCTATCGACGGCGCTATGGTTCGGTTCGGGCTAT 68124  
 665 AspGlnSerSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgAr 681  
 68125 GATCAATCCTCTTTGGGTCCACGCTCACAAGCCTATTTGACAGCTCGCTG 68174  
 681 gGlyGlnGlnThrThrLeuGlyGluValValIcIyGlyAsnAlaLeuAlaT 698  
 68175 TGGTCAACAAACCACTAGGAGAGGTTGTTGGTGGTAAATGCTTTGGCAA 68224  
 698 hrPheGlySerGluLeuIleLeuProLeuProPheLysGlyAspTriPle 714  
 68225 CTTTCGGCAGTACGCTGATTTTACTTTTGCCTTTAAAGGTGATTGGATA 68274  
 715 AspGlnValArgProValIlePheIleGluGlyGlnValPheAspTh 731  
 68275 GATCAGGTGCGTCCAGTGAATATTCATTGAGGCGGTCAGGTTTTTGATAC 68324  
 731 rThrGlyMetAspLysGlnThrIleAspLeuThrGlnPheLysAspProG 748  
 68325 AACAGGTATGGATAACAACAACCATTTGATTTAACCCCAATTTAAAGACCAC 68374

748 lnaLaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuThr 764  
|||||  
68375 AAGCAACAGCTGAACAAATGCAAAAGCAGCCAAATCGCCGCTACTAAAC 68424  
|||||  
765 GlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrTrpTyrTh 781  
|||||  
68425 CAAGATAAACAGTGGCTTATAGTGTGGTGGTGGCAACTGGGTATAC 68474  
|||||  
781 rProLeGlyProLeuSerIleSerTyrAlaLysProLeuAsnLysLysG 798  
|||||  
68475 GCCCATGTGCTTATATCTATAGCTATGCCAAGCCATTTGAATAAAAC 68524  
|||||  
798 lnaAsnAspGlnThrAspThrValGlnPheGlnIleGlySerValPhe 813  
|||||  
68525 AAAATGATCAGACCGATACGATCGGTACAGTTCAGATTGGTAGTGCTTT 68571  
|||||

seq\_name: gb\_pat:AX081536

seq\_documentation\_block:  
LOCUS AX081536 1000 bp DNA linear PAT 27-FEB-2001  
DEFINITION Sequence 41 from Patent WO0109350.

ACCESSION AX081536  
VERSION AX081536.1 GI:13170355

KEYWORDS  
SOURCE Moraxella catarrhalis.

ORGANISM Moraxella catarrhalis  
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;

Moraxella.

REFERENCE 1 (bases 1 to 1000)

AUTHORS Berthet,F.X., Dalemans,W.L., Denoel,P., Dequesne,G.S., Peron,C.S.,  
Lobet,Y.S., Poolman,J.S., Thiry,G.S., Lhonnard,J.S. and Voet,P.S.

TITLE Genetically engineered bleb vaccine

JOURNAL Patent: WO 0109350-A 41 08-FEB-2001;

SMITHKLINE BEECHAM BIOLOGICALS S.A. (BE)

FEATURES  
Location/Qualifiers

source

1..1000  
/organism="Moraxella catarrhalis"

/db\_xref="taxon:480"

BASE COUNT 287 a 172 c 215 g 326 t

ORIGIN

alignment\_scores:  
Quality: 36.00 Length: 36  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-701-711-2 x AX081536 ..

Align seg 1/1 to: AX081536 from: 1 to: 1000

1 MetArgAsnSerTyrPheLysGlyPheGlnValSerAlaMetThrMetal 17  
|||||  
821 ATGCGTAATTCATATTTTAAAGGTTTCAGTTCAGTGCATGACATGGC 870  
|||||

17 aValMetMetValMetSerThrHisAlaGlnAlaAlaAspPheMetalAa 34  
|||||

871 TGTCATGATGTAATGCACTCATGCAACAGCGCGGATTTTATGGCAA 920  
|||||

34 snAspIle 36

|||||

921 ATGACATT 928

seq\_name: gb\_pat:AX009293

seq\_documentation\_block:

LOCUS AX009293 61 bp DNA linear PAT 06-SEP-2000

DEFINITION Sequence 7 from Patent WO9963093.

ACCESSION AX009293

VERSION AX009293.1 GI:9996622

KEYWORDS

SOURCE synthetic construct.

ORGANISM synthetic construct  
artificial sequence.  
REFERENCE 1 (bases 1 to 61)  
AUTHORS Vinals-Bassols,C.  
TITLE Basb027 proteins and genes from moraxella catarrhalis, antigens,  
antibodies, and uses  
JOURNAL Patent: WO 9963093-A 7 09-DEC-1999;  
SMITHKLINE BEECHAM BIOLOG (BE); VINALS BASSOLS CARLOTA (BE)

FEATURES  
Location/Qualifiers

source

1..61  
/organism="synthetic construct"

/db\_xref="taxon:32630"

/note="Oligonucleotide"

BASE COUNT 20 a 15 c 17 g 9 t

ORIGIN

alignment\_scores:

Quality: 11.00 Length: 11  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-701-711-2 x AX009293 ..

Align seg 1/1 to: AX009293 from: 1 to: 61

39 ThrGlyLeuGlnArgValThrIleGluSerLeu 49  
|||||

28 ACAGGACTACAGCGAGTGACCATTTGAAAGCTTA 60  
|||||

seq\_name: gb\_pat:AX009294

seq\_documentation\_block:

LOCUS AX009294 67 bp DNA linear PAT 06-SEP-2000

DEFINITION Sequence 8 from Patent WO9963093.

ACCESSION AX009294

VERSION AX009294.1 GI:9996623

KEYWORDS

SOURCE synthetic construct.

synthetic construct.  
artificial sequence.

REFERENCE 1 (bases 1 to 67)

AUTHORS Vinals-Bassols,C.

TITLE Basb027 proteins and genes from moraxella catarrhalis, antigens,  
antibodies, and uses

JOURNAL Patent: WO 9963093-A 8 09-DEC-1999;

SMITHKLINE BEECHAM BIOLOG (BE); VINALS BASSOLS CARLOTA (BE)

FEATURES  
Location/Qualifiers

source

1..67  
/organism="synthetic construct"

/db\_xref="taxon:32630"

/note="Oligonucleotide"

BASE COUNT 22 a 16 c 15 g 14 t

ORIGIN

alignment\_scores:

Quality: 11.00 Length: 11  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-701-711-2 x AX009294/rev ..

Align seg 1/1 to reverse of: AX009294 from: 1 to: 67

803 AspThrValGlnPheGlnIleGlySerValPhe 813  
|||||

66 GATACGGTACAGTTCCAGATTGGTAGTGCTCTT 34  
|||||

seq\_name: gb\_ba:PL0236920



```

seq_documentation_block:
LOCUS      PLU236920               4405 bp    DNA          linear      BCT 02-AUG-1999
DEFINITION Photorhabdus luminescens yael (partial), fira (partial), oma and ompH genes.
ACCESSION  AJ236920
VERSION    AJ236920.1 GI:5689864
KEYWORDS   fira gene; oma gene; ompH gene; outer membrane antigen; periplasmic protein; yael gene.
SOURCE     Photorhabdus luminescens.
ORGANISM   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Photorhabdus.
REFERENCE  1 (bases 1 to 4405)
AUTHORS    Chatonnet-Marton,P.I., Givaudan,A., Lanois,A. and Boemare,N.E.
TITLE      Photorhabdus luminescens genomic region homologous to 4.0 minute Escherichia coli region promotes pleiotropic phenotypes
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 4405)
AUTHORS    Chatonnet-Marton,P.I.
TITLE      Direct Submission
JOURNAL    Submitted (12-FEB-1999) Chatonnet-Marton P.I., Laboratoire de Pathologie Comparee, Institut National de la Recherche Agronomique, Universite Montpellier II, C.C.101, Pl. E. Bataillon, Montpellier, 34095, FRANCE

FEATURES             Location/Qualifiers
     source            1..4405
                     /organism="Photorhabdus luminescens"
                     /strain="Hm"
                     /db_xref="taxon:29488"
     gene              1..681
                     /dev_stage="phase I variant"
                     /genes="yael"
     CDS                1..681
                     /genes="yael"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB51928.1"
                     /db_xref="GI:5689865"
                     /translation="MEKVTPGSAAEKAGKQGDRIYKVGSOEIDVHMTFTSFVSNPN
                     VPLESDVRAGHLISLMTPEYQSGGRKGVGFAGVELRIVPLADEYKIVQOYGPESA
                     MYQAGDTPQLMRLTVSMIGKLIVGDVTKVNNLSIGAKGAGSADSLGVYLLMFLA
                     LKVNGLINLPLPVLDDGHLFLFIEIKIGPVSERVQDFSYRGMILVLMGLA
                     LFNDSRF"
     RBS                696..703
                     /genes="oma"
     gene              join(696..703,716..3109)
                     /genes="oma"
     CDS                716..3109
                     /genes="oma"
                     /note="putative"
                     /note_start=1
                     /transl_table=11
                     /product="outer membrane antigen"
                     /protein_id="CAB51929.1"
                     /db_xref="GI:5689866"
                     /translation="WAMKKLLIASLLFGSAAAYGADGFVQDHFEGLRQVAVGAALL
                     NNPVRVQDVEDIGRTTHALFATGNFEDVRVRDNGTLIVQVKERTPIASITFSGN
                     KSKVDMLQKNGLEASHVRVGEALDRTMLNIEGLEDEYVSVGNASVAVVTPLR
                     NRVDLKLVFAGVSAKIQINIVGNKSFSSDELLNRFLQRDRDDVPWNLTADQYKQK
                     LGDLEALRSFYLDGRYAFNRINDQVSLTPDKGIYVITNNTEGDQYKISIDLNGN
                     MAGYQETFKLAAIEPGLSYNGTQVTKMENDIKNLGRYGYAYPRVMTQPEINDQDKT
                     LKLVHINDAGNRVYVRKIRFSGNDITKDSVLRRMRQMERAWLSDVLBELGERLNRLL
                     GYFETVDVETQIRPGSPQDVVYVKVKERNTGSLNFGVGFGEVSGVFOIGAOQDNL
                     CTGNVAGINASKNDVSTYAELSFDPYFTINGVSLGGRVYNDFRADDAELSGYTNOS
                     YGISGLGFPFINENSLNGLNIIHNSLDMLPQVAMWRYLRSMGEKPDLSKAFEKA
                     DFLALPMGTGINLDRGFPTSGVSTLNGKVTIPGSDFNEFYKVTLDYSAYYIINDRR
                     TWILGRSLRGYDGLGKELPEFYENFYAGTSFRSNNIGKAIYLYKDGSPKK
                     EPSRDAVGCGNAMAVALSLITPTPLFDKYSNSVTSFIDSGTWMTDWDSAVMK
                     SKGIPDYSKPGNIRYSAGIALQWMSPLGLPVSYAKPIKDYEGDRSEQFPFNGTWT"
     gene              3055..3070
                     /genes="ompH"
     -35_signal        3065..3070
                     /genes="ompH"

```

```

seq_documentation_block:
LOCUS      PLU236920               4405 bp    DNA          linear      BCT 02-AUG-1999
DEFINITION Photorhabdus luminescens yael (partial), fira (partial), oma and ompH genes.
ACCESSION  AJ236920
VERSION    AJ236920.1 GI:5689864
KEYWORDS   fira gene; oma gene; ompH gene; outer membrane antigen; periplasmic protein; yael gene.
SOURCE     Photorhabdus luminescens.
ORGANISM   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Photorhabdus.
REFERENCE  1 (bases 1 to 4405)
AUTHORS    Chatonnet-Marton,P.I., Givaudan,A., Lanois,A. and Boemare,N.E.
TITLE      Photorhabdus luminescens genomic region homologous to 4.0 minute Escherichia coli region promotes pleiotropic phenotypes
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 4405)
AUTHORS    Chatonnet-Marton,P.I.
TITLE      Direct Submission
JOURNAL    Submitted (12-FEB-1999) Chatonnet-Marton P.I., Laboratoire de Pathologie Comparee, Institut National de la Recherche Agronomique, Universite Montpellier II, C.C.101, Pl. E. Bataillon, Montpellier, 34095, FRANCE

FEATURES             Location/Qualifiers
     source            1..4405
                     /organism="Photorhabdus luminescens"
                     /strain="Hm"
                     /db_xref="taxon:29488"
     gene              1..681
                     /dev_stage="phase I variant"
                     /genes="yael"
     CDS                1..681
                     /genes="yael"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB51928.1"
                     /db_xref="GI:5689865"
                     /translation="MEKVTPGSAAEKAGKQGDRIYKVGSOEIDVHMTFTSFVSNPN
VPLESDVRAGHLISLSMTPVQSGGRKVGFGAGVELRIVPLADEYKIVQOYGPESA
MYQAGDTPQLMRLTVSMIGKLLIVGDVTKNNLSPISTAKGAGVSADSGLLVYLMLFA
LISVNGIINLIPLPVLDGGHLLFLFIEIKIGPVSERVQDFSYRGMILVLMGLA
LNFDSRF"
     RBS                696..703
                     /genes="oma"
     gene              join(696..703,716..3109)
                     /genes="oma"
     CDS                716..3109
                     /genes="oma"
                     /note="putative"
                     /note_start=1
                     /transl_table=11
                     /product="outer membrane antigen"
                     /protein_id="CAB51929.1"
                     /db_xref="GI:5689866"
                     /translation="WAMKKLLIASLLFGSAAAYGADGFVQDIHFEGLRQVAVGAALL
NMPVRVQDVSDEIDGTTHALFATGNFEDVRVRDNGTLIVQVKERTPIASITFSGN
KSKVDMLKQNLGEALSHVRGDELTMLNIEGLEDEYVSVKYNASKAVVTPLR
NRVDLKLVFAEVSASAKIQINIVGNKSFSSDELLNRFLQRDRDDVPWNLTDQKYOKQK
LGDLEALSRYFLDRGARFNIDQVSLTPDKGIYVTINMTEGDQYKISIDLNGN
MAGYQETFKLAAIEPGLSYNGTQVTMENDIKNLGRYGAYPRVMTQPEINDQDKT
LVKLVINDAGNRVYVRKIRFSGNDITKDSVLRRMRQMERAWLSDVLBELGERLNRLL
GYFTVDVETQRIPGSPQDVVYVKKERNTGSLNFGVGFGEVSGVFOIGAOQDNRL
GTGNVAGINASKNDVSTYAELSFDPYFTINGVSLGGRVYNDFRADDAELSGYTNOS
YGISGLGFPFINENSLNGLNLTIIHNSLDMLPQVAMWRYLRSMGEKPDLSKAFEKA
DFAFLPMGTGINLNDRGFTTSGVSTLNGKVTIPGSDFNEFYKVTLDYSAYYIINDRR
TWILGRSLRGYDGLGKELPEFYENFYAGTSFRSNNIGKAIYLYKDGSPKK
EPSRDAVGGNAMAVASLELITPTPLDYSKYSNVRTSPTIDSGTWDTDMWDSAVMK
SKGIPDYSKPGNIRYSAGIALQWMSPLGLPVSYAKPIKDYEGDRSEQFPFNGTWT"
     /genes="oma"
     /genes="ompH"
     /genes="ompH"

-10_signal
gene
gene
RBS
CDS

BASE COUNT      1187 a      819 c      1094 g      1305 t
ORIGIN

alignment_scores:
Quality:      11.00      Length:      11
Ratio:        1.000      Gaps:        0
Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x PLU236920      ..

Align seg 1/1 to: PLU236920 from: 1 to: 4405

647  LeuProPheTyrGluAsnPheTyrAlaGlyGly  657
|||||
2648  TTGCCATTCTATGAAAACTTCTATGCAGCGGC  2680

seq_name: gb_pat:AX078498

seq_documentation_block:
LOCUS      AX078498               1830 bp    DNA          linear      PAT 22-FEB-2001
DEFINITION Sequence 12 from Patent WO0107624.
ACCESSION  AX078498
VERSION    AX078498.1 GI:13158140
KEYWORDS   Pseudomonas putida.
SOURCE     Pseudomonas putida
ORGANISM   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; Pseudomonas.
REFERENCE  1 (bases 1 to 1830)
AUTHORS    Fraser,C.M., Venter,C., Tuemmler,B., Hoheisel,J., Duesterhoeft,A., Hilbert,H., Timmis,K.N., Moore,E., Straetz,M., Helm,S.,

```

Nelson, K.E., Hickey, E. and Peterson, J.  
 Dna sequences which are suited for specifically detecting  
 pseudomonas putida kt2440  
 Patent: WO 0107624-A 12 01-FEB-2001;  
 THE INSTITUTE FOR GENOMIC RESEARCH (US) ; OIAGEN GmbH (DE) ;  
 Gesellschaft fuer Biotechnologische Forschung mbH (GBF) (DE) ;  
 Deutsches Krebsforschungszentrum (DKFZ) (DE) ; Medizinische  
 Hochschule Hannover (DE)  
 FEATURES  
 source Location/Qualifiers  
 1..1830  
 /organism="Pseudomonas putida"  
 /db\_xref="taxon:303"  
 BASE COUNT 401 a 555 c 529 g 344 t 1 others  
 ORIGIN

alignment\_scores:  
 Quality: 10.00 Length: 10  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-701-711-2 x AX078498 ..  
 Align seg 1/1 to: AX078498 from: 1 to: 1830  
 368 A\$PGLuValLeuArgGluMetArgGln 377  
 |||||  
 1081 GAGCAAGTGTGCGTGGCGAATGCCCGAG 1110

seq\_name: gb\_ba:AF120927

seq\_documentation\_block:  
 LOCUS AF120927 2433 bp DNA linear BCT 02-MAR-2001  
 DEFINITION Shigella flexneri outer membrane antigen Oma90 (oma90) gene,  
 complete cds.  
 ACCESSION AF120927  
 VERSION AF120927.1 GI:4567045  
 KEYWORDS  
 SOURCE Shigella flexneri.  
 ORGANISM Shigella flexneri.  
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 Shigella.

REFERENCE  
 1 (bases 1 to 2433)  
 Robb, C.W., Orihuela, C.J., Ekkelenkamp, M.B. and Niesel, D.W.  
 Identification and characterization of an in vivo regulated  
 D15/Oma87 homologue in Shigella flexneri using differential display  
 polymerase chain reaction  
 JOURNAL Gene 262 (1-2), 169-177 (2001)  
 PUBMED 11179681

REFERENCE  
 2 (bases 1 to 2433)  
 Robb, C.W., Ekkelenkamp, M.B., Orihuela, C.J. and Niesel, D.W.  
 Direct Submission  
 TITLE Submitted (16-JAN-1999) Microbiology and Immunology, University of  
 Texas Medical Branch, 301 Administration Blvd, Galveston, TX 77555,  
 USA

FEATURES  
 source Location/Qualifiers  
 1..2433  
 /organism="Shigella flexneri"  
 /db\_xref="taxon:623"  
 1..2433  
 /gene="oma90"  
 1..2433  
 /gene="oma90"  
 /note="similar to Escherichia coli hypothetical YaeT, 90  
 kDa"  
 /codon\_start=1  
 /transl\_table=11  
 /product="outer membrane antigen Oma90"  
 /protein\_id="AAD23568.1"  
 /db\_xref="GI:4567046"  
 /translation="MAMKLLIASLFFSSATVYAGEGFVVKDIIHFEGLQRVAVGAALL  
 SMPVRTGDTVNDISNTIRALFATGNFEDVRLRDGDTLLVQVKERTIASITFSGN

KSVKDDMLKQNLASGYRVGSLDRTTIADIEKLEDFYISVGKYSASVKAVVPLPR  
 NRVDKLVFQEGVSAETQQINIVGNHAFITDDLSHFQRLDEVWNNVVGDRKKQKQK  
 LAGDLETLRYLDRYARFNIDTSQVSLTPDKKGIYVTVNITGDDQTKLSGVEVSGK  
 LAGHSAETQLTKIEPGLYNGTKVTKMEDDIKKLGRYGVAYPRVQSMPEINDADKT  
 VKLRVNDAGNRVFRKFRFEGNTSKDAVLRREMROGAWLGSIDLVDQGRKLRNL  
 GFEFTVDTRQVPSDPQDVVYKVERNTGSFNGIGYCTESGVSFQAGVQODNWL  
 GTGVAVGINGTKNDYQYIAELSVINPFTVDCVSLGRLFYNDQFQADADLSDYTNKS  
 YGTDVTIGFPINEYNSRAGLGYVHNSLNNQPVAMWRYLISGHEHPTSDQDN\$FK  
 TDDTFNYGWTYINKLDRGYFPTDGSRVNLTKVTIPGSDNEYIKVTLDTATYVPIDDD  
 HKWVVLGTRWYGDGLGKEMPFYFAGGSSTVRGFSQNTIGPKAVYFPHQASNY  
 DPDYDECATODGAKDLCKSDDAVGGNMAVASLEFIPPTPI\$DKYANSVRTSFFWD  
 MGTWMDTWNDSOYSGYPDYSNIRMSAGIALQWNSPLGLPVSQAQPKKYDGDK  
 AEQFQPNIGTKW"  
 BASE COUNT 606 a 586 c 647 g 594 t  
 ORIGIN

alignment\_scores:  
 Quality: 10.00 Length: 10  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-701-711-2 x AF120927 ..  
 Align seg 1/1 to: AF120927 from: 1 to: 2433  
 648 ProPheTyrGluAsnPheTyrAlaGlyGly 657  
 |||||  
 1939 CCGTCTCAGAGAACTTCTATGCCGGTGT 1968

seq\_name: gb\_ba:AF407013

seq\_documentation\_block:  
 LOCUS AF407013 2700 bp DNA linear BCT 10-SEP-2001  
 DEFINITION Escherichia coli outer membrane protein (ecfK) gene, complete cds.  
 ACCESSION AF407013  
 VERSION AF407013.1 GI:15529631  
 KEYWORDS  
 SOURCE Escherichia coli.  
 ORGANISM Escherichia coli.  
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 Escherichia.

REFERENCE  
 1 (bases 1 to 2700)  
 Dartigalongue, C., Missiakas, D. and Raina, S.  
 Characterization of the Escherichia coli sigma E regulon  
 J. Biol. Chem. 276 (24), 20866-20875 (2001)  
 MEDLINE 21293005  
 PUBMED 11274153

REFERENCE  
 2 (bases 1 to 2700)  
 Dartigalongue, C., Missiakas, D. and Raina, S.  
 Direct Submission  
 TITLE Submitted (09-AUG-2001) Biochimie Medicale, CMU, 1 Rue Michel  
 Servet, Geneva 1211, Switzerland  
 FEATURES  
 source Location/Qualifiers  
 1..2700  
 /organism="Escherichia coli"  
 /db\_xref="taxon:562"  
 61..>2640  
 /gene="ecfK"  
 61..100  
 /gene="ecfK"  
 /note="factor sigma24"  
 101..>2640  
 /gene="ecfK"  
 /note="outer membrane protein"  
 208..2640  
 /gene="ecfK"  
 /note="Ecfc; under heat shock transcriptional control of  
 RpoE; not essential for Escherichia coli growth; similar  
 to Oma87 and D-15 antigen"  
 /codon\_start=1  
 /transl\_table=11

```
/product="outer membrane protein"
/protein_id="AAL01379.1"
/db_xref="GI:1529632"
/translation="MAKKLLIASLLFSSATVYGAEGFVVVDIHFEGLOQVAVGAALL
SMFVTCGTVDNDEISNTIRALFATGNFEDVRVLRDGTLLVQVKERPTTASTITFSGN
KSVKDKMLKNLEASGVYRGESLDRFTIADIEKGLDFYYSVGKYSASVKAATVPLPR
NRVDLKLFOEGYSAEIQINIVGNHAFITDELISHFOLRDEVPWNVGVGDRYQOK
LAGDLFRLSYILDRGYARENDISTOVSLTPDKKGIYVTNITEGDOYKLSGVEVSGN
LAGSAIEQLTKIEGELYNGTKYKMEDDIKLLRGYGYAYPRVQSHPEINDAKT
VKLUNVNDAGNRVYVRKIRFEGNDTSKDAVLRMRQMEGAWLGSLDVGOKERLNL
GFETVDTQRPVGSQDVVYKVKERTGTSFNFGIGYGTSGVSYFQAGVQODNML
GTGAVGINTKNDQTYAELSVNTNFTYVDGVSGLGRFLYNDFOADDLSDITNKS
YGVDTVLGFPINEYNSLRAGLVHNSLNMQPOVAMRWLYL SMGHPSTSDQDNSFK
TDDFTNGYTYNKLDRGYEPTDGSRVNLTKVTIPGSDNEYKVTLDATATYVPIDD
HKWYLGRTRMGVDGLGKEMPEFYENYAGGSSVTRGFSQNTIGPKANYFPHQASNY
DPDIECATQDGAOKDLCKSDDAVGGNAMVASLEFIPTPFIISDKYANSVTSFWD
MGVWTDNTSSQSYGYPDYSPTNSIRMSAGIALQWNSPLGLVFSYAQPFKKYDGGK
AEQFQFNIGKTW"
BASE COUNT      668 a      636 c      721 g      675 t
ORIGIN

alignment_scores:
  Quality:      10.00      Length:      10
  Ratio:        1.000      Gaps:      0
  Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
  US-09-701-711-2 x AF407013      ..

  Align seg 1/1 to: AF407013 from: 1 to: 2700

      648  ProPheTyRGluAspPheTyRAlaGLyGLy 657
      2146  CGGTCTACGAGAACTTATGTCGCGTGGT 2175

seq_name: gb_ba:AE005193

seq_documentation_block:
LOCUS      AE005193
DEFINITION      Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 17
of 155.
ACCESSION      AE005193 AE005174
VERSION        AE005193.1 GI:12512893
KEYWORDS
SOURCE        Escherichia coli O157:H7 EDL933.
  Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
  Escherichia.
REFERENCE      1 (bases 1 to 10649)
  Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
  Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
  Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
  Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamoudis,K.,
  Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
  Welch,R.A. and Blattner,F.R.
  Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
  Nature 409 (5819), 529-533 (2001)
  21074935
  11206551
REFERENCE      2 (bases 1 to 10649)
  Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
  Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
  Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
  Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamoudis,K.,
  Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
  Welch,R.A. and Blattner,F.R.
  Direct Submission
  Submitted (22-OCT-2000) Laboratory of Genetics, University of
  Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
  Location/Qualifiers
    1..10649
    /organism="Escherichia coli O157:H7 EDL933"

/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohemorrhagic"
215..940
/gene="rpsB"
/note="20180"
215..940
/gene="rpsB"
/function="structural component; Macromolecule synthesis,
modification: Ribosomal proteins - synthesis,
modification:riboso"
/note="Residues 1 to 241 of 241 are 100.00 pct identical
to residues 1 to 241 of 241 from Escherichia coli K-12
Strain MG1655; B0169"
/codon_start=1
/transl_table=11
/product="30S ribosomal subunit protein S2"
/protein_id="AAG54471.1"
/db_xref="GI:12512894"
/translation="MATVSMRDLKAGVHFGHQTQRYWPKMKPFIFGARNKVHIINLE
KTYPMNEALAEALNKIASRKGKILFVCTKRAASEAVKDAALSCDQFFVNHFWLGMILT
NWKTVROSIRKLKDLQTSQDGTDFKLTKEALMRTRELEKLENSLGKIDKMGSLPDA
LFVIDADHEHIAIKEANLGIPIVFAIVDTNSDPDGVDFVIPGNDDAIRAVTLYLGAVA
ATVREGRSODLASQAEESEVEAE"
1198..2049
/gene="tsf"
/note="20181"
1198..2049
/gene="tsf"
/function="factor; Macromolecule synthesis, modification:
proteins - translation and modification"
/note="Residues 1 to 283 of 283 are 100.00 pct identical
to residues 1 to 283 of 283 from Escherichia coli K-12
Strain MG1655; B0170"
/codon_start=1
/transl_table=11
/product="protein chain elongation factor EF-Ts"
/protein_id="AAG54472.1"
/db_xref="GI:12512895"
/translation="MAEITASLVKELRERTGAGMDCCKKALTEANGDIELAIENWRKS
GATKAAGKAGNVAADGVKTKIDNGYIILEVNCQTFVAKDGFQAFADKVLDAAGK
GKITDVELKAQFEERVALVAKIGENINIRVAALLEGVDLGSYOHGARGVLVAAGK
ADELKVHIAMHVAASKPEFTKPEDVSAEVVEYQVOLDIAMQSGPKKEIAERWEG
RMKKTGEVSLTGQPFVMEPSKTVGQLKKEHNAEVTGFIREFGEGIEKVTDTFAAEV
AAMSKQS"
2196..2921
/gene="pyrH"
/note="20182"
2196..2921
/gene="pyrH"
/function="enzyme; Central intermediary metabolism:
Nucleotide interconversions"
/note="Residues 1 to 241 of 241 are 100.00 pct identical
to residues 1 to 241 of 241 from Escherichia coli K-12
Strain MG1655; B0171"
/codon_start=1
/transl_table=11
/product="uridylylate kinase"
/protein_id="AAG54473.1"
/db_xref="GI:12512896"
/translation="MATNAKPVYKRILLKLSGALQGTGEGIDASILDRMAQEIKEK
VELGIQGVVIGGNLFRGAGLAKAGNRVVGDMGLATVNGGLAMRDALHRAVVA
RLMSAIPLNGVCSYSWAEALSLRNRRVILSAGTGNPFETDSACLRCGIEADY
VMKPKVDGVFTADPAKDPATMYEQITYSEVLEKELKVMDLAAFTLARDHKLPIRVF
LNKPKGALRRVYMGKEGEGTLITE"
3071..3628
/gene="frr"
/note="20183"
3071..3628
/gene="frr"
/function="factor; Macromolecule synthesis, modification:
proteins - translation and modification"
```

/note="Residues 1 to 185 of 185 are 100.00 pct identical to residues 1 to 185 of 185 from Escherichia coli K-12 Strain MG1655: B0172"

gene

CDS

/transl\_table=11

/product="ribosome releasing factor"

/protein\_id="AAG54474.1"

/db\_xref="GI:12512897"

/translation="MISDIRKDAEVRMDKCVFAFKTIQISKIRTFGRASPSLLDGIIVVEY YGTPPLQLASVTVEDSRTLKINVFDRSMSPAVEKAIMASDLGNPNAGSDIRVPL PPLTERKDKLTKIVRGAEOARVAVRNVRDANDKVKALKDKKEISDDDDRRSDDDV OKLTDAAIKKIEAALADKAEALMQF"

3720. .4916

/gene="yaem"

/note="20184"

3720. .4916

/gene="yaem"

/function="putative transport; Not classified"

/note="Residues 1 to 398 of 398 are 99.74 pct identical to residues 1 to 398 of 398 from Escherichia coli K-12 Strain MG1655: B0173"

/codon\_start=1

/transl\_table=11

/product="putative ATP-binding component of a transport system"

/protein\_id="AAG54475.1"

/db\_xref="GI:12512898"

/translation="MKQTLTGSTGICGCTLDVVVRHNPFRVVALVAGKNVTRMVE OCLEFSPYAVMDDEASAKLLKMLQCGSRTEVLSSGQACDMALEDVQVMAIV GAAGLLPTLAIRAGTKILLANKESLVTGRLFMADVAKQSKAQLLPVPSEHNAIQSL PQPIQHNGLYADLEONGVSVILLTGSGGPFRETPLRDLATMPDQACRHPNWSMGRKI SVDSATMKNKGLLEYEARLWFNASQOMEVLHPQSVIHSVMRYQDGSVLQAOLGEPDM RPTIAHTMAWPNRVNSGVKPLDFCKLSALTAAAPDYDRYPCCLKLAMEAFEGQQAATTA LNAANEITVAFAFQAQRIREFTDIAALNLSVLEKMDMREPQCVDVLSVDASAREVARKE VMLAS"

5102. .5863

/gene="yaes"

/note="20185"

5102. .5863

/gene="yaes"

/function="orf; Unknown function"

/note="Residues 1 to 253 of 253 are 100.00 pct identical to residues 1 to 253 of 253 from Escherichia coli K-12 Strain MG1655: B0174"

/codon\_start=1

/transl\_table=11

/product="orf, hypothetical protein"

/protein\_id="AAG54476.1"

/db\_xref="GI:12512899"

/translation="MMLSATQPLSEKLPAGCRHVVAIIMDNGRWAKQKIRAFGHK AGAKSVRAVSFAANNQIEALTLYAFSSNNRPAQEVSAIMELFWALDSEVKSLSHR HNVRLRIIGDTSRFRNSRIQERIRKSEALTAGNTGLTLNIAANYGRMDIVQGVROLAE KVOQNLQPDQIDEMLNOHVMHAPDLVIRVTGGEHRI.SNELLWQIAYAELYFTD VLPDPDFEQDFEGALNAFANRRERFGGTEPGDETA"

5984. .6733

/gene="cdsa"

/note="20186"

5984. .6733

/gene="cdsa"

/function="enzyme; Fatty acid biosynthesis: Fatty acid and phosphatidic acid biosynth"

/note="Residues 1 to 249 of 249 are 100.00 pct identical to residues 1 to 249 of 249 from Escherichia coli K-12 Strain MG1655: B0175"

/codon\_start=1

/transl\_table=11

/product="CDP-diglyceride synthetase"

/protein\_id="AAG54477.1"

/db\_xref="GI:12512900"

/translation="MLAAEWGQSLSGFTTSRQVRVLAIVLCGLLLALMLFLPEYHRNI HOPVLEISLAWGLWIVALLLVLFYPGSAAIWRNSKTLRIFGLVITVPPFWGLAL RAWHYDENYSGATWLLYVILLVMGADSGAYMFGKLFKGLKHLAPKVPKPGKQFIFG LATAVISMGYGMWANDVAPVTLICSVIAALASVLGDLTSMFKREAGIKDSGLHI

POHGILDRIDSLTAAPVPVFACLLLLVFRFL"

6745. .8097

/gene="yael"

/note="20187"

6745. .8097

/gene="yael"

/function="orf; Unknown function"

/note="Residues 1 to 450 of 450 are 100.00 pct identical to residues 1 to 450 of 450 from Escherichia coli K-12 strain MG1655: B0176"

/codon\_start=1

/transl\_table=11

/product="orf, hypothetical protein"

/protein\_id="AAG54478.1"

/db\_xref="GI:12512901"

/translation="MLSFWDIASFTVALGVLTIVHEFGHFWARRCGVVERFSTIGF GKALWRTDKLGTETVYVIALPLGGYVKMLDERAEPVVPRLRHHFNKSVGQRAAIIA AGPVANFIFAIFFAYLVLFVIGVPGVPVVEGAANSAIAEAQIAPGTDLKAVDGIETP DWDAYRLQLVDKIGDESTITVAPFGSQQRDRVKLDLRHWAPEPKDPPVSLGIRPR GPQIEPVLENVQPNASAKAGLQAGDRIVKVDGQPLTQWVTFVMLVRDNPGRKSLALEI

alignment\_scores:

Quality: 10.00 Length: 10

Ratio: 1.000 Gaps: 0

Percent similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-701-711-2 x AE005193

Align seg 1/1 to: AE005193 from: 1 to: 10649

648 ProPheTyrGluAsnPheTyrAlaGlyGly 657

|||||

10065 CCGTCTACGAGAACTTCTATGCCGGTGT 10094

seq\_name: gb\_ba:AE004297

seq\_documentation\_block:

LOCUS AE004297

DEFINITION Vibrio cholerae chromosome I, section 205 of 251 of the complete chromosome.

ACCESSION AE004297 AE003852

VERSION AE004297.1 GI:9656810

KEYWORDS

SOURCE Vibrio cholerae.

ORGANISM Vibrio cholerae

REFERENCE

1 (bases 1 to 10901)

Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickley, E.K., Peterson, J.D., Umayam, L., Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D., Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I., Sellers, P., McDonald, L., Utterback, T., Fleischmann, R.D., Nierman, W.C. and White, O.

DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae

Nature 406 (6795), 477-483 (2000)

20406833

REFERENCE

2 (bases 1 to 10901)

Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickley, E.K., Peterson, J.D., Umayam, L., Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D., Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I., Sellers, P., McDonald, L., Utterback, T., Fleischmann, R.D., Nierman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R., Mekalanos, J.J., Venter, J.C. and Fraser, C.M.

Direct Submission

Submitted (14-JUN-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

Location/Qualifiers

1. .10901

/organism="Vibrio cholerae"

/strain="N16961"

```
/serotype="O1"
/db_xref="taxon:666"
/chromosome="I"
/notes="biotype: El Tor"
/complement(69. .1124)
/gene="VC2250"
/complement(69. .1124)
/gene="VC2250"
/notes="similar to GB:X54797 SP:P21645 PID:1208950
PID:41470 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="UDP-3-O-3-hydroxymyristoyl glucosamine
N-acyltransferase"
/protein_id="AAF95394.1"
/db_xref="GI:9656811"
/translation="MTTLTLELATITGGELFGDPALVSAPMPKQALGHVTFILSN
PKYSKHLGDKATVIMKTSERELCPNSALVADPYAFKVAQALDSTPSPAHGIAP
SAVIAEDAKLGHNVISGANAVIESGVQLGDNVIGAGCFIGKQARLGNTPKLANVTI
YHKVEIGSDCLIQSGTVIGADGFGYANERGEWIKIPQLGSVRIGDRVEIGACTTIDRG
ALDDTVIEDNVIIDNLOLTAHNVHIGYSGALAGCTVIAGSTRIGKYCIIGASVINGH
IETADGVITIGMGWRSIEEKGMYSGGIPLQTNKEWRKTAARVHRIDDMHKKLKALE
KLEQSDTVQPDNSQAE"
/complement(1134. .1643)
/gene="VC2251"
/complement(1134. .1643)
/gene="VC2251"
/notes="similar to PID:1935023; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="outer membrane protein OmpH"
/protein_id="AAF95395.1"
/db_xref="GI:9656812"
/translation="MKMKIAASLGILIISSSMANABAAQKIGVINTAQVFOALPQ
REVVLQKMEEFKDKAAELQAIQADAKRIEKLROGLMQDQDEVEKRIETIGQLDSK
YKIAQALQASARAEAEKQLFKVIQDAVKVAKEGYDIVLDTSSMQYGRKEPHNL
SEKVIKAIK"
/complement(1658. .4069)
/gene="VC2252"
/complement(1658. .4069)
/gene="VC2252"
/notes="similar to GB:U00096 SP:P39170 PID:1552754
PID:1786374; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="surface antigen"
/protein_id="AAF95396.1"
/db_xref="GI:9656813"
/translation="MAMKQILLATLATSANGAEKFFVQDQIDGLORVALGAALL
KMPVRVGSVDQVANIILKALYSSGNFEDVKVLRDGNLTMVQKERPTIASVSFSGN
KAKEBQLQKNEASIRVEGALDRITLSEKLEDFIYSVGKYNATKAVVYPLPR
NRADLKFTVEGSAKIQIINFIGNQVDSDELLSFNLDNVAMWNLADDKYKQV
LAGDIBALTYLDRLGYLFQVDSTQVAISPDKKGVYITLNLNEGSPYTVKVPFRGE
LMGKEAFTSLIPFEIGETNGSAVTRLEESKVLGESCAYAPVORTIPEFDDEKQO
VSLVHVHAGKRVYVDIRFVGNNSRDEVLREMQEGSWLNSKDIETGKTRNLRL
GFTEFVEQTVRVPGESQVDLYVSKEANSNVNFGVIGTESGVSFQVGLQDNNFL
GSGERVGVAMINDYOKNLTRDYPNLDGVSGLGVKVFYQNFSEAGIYDNTNES
YDTSITWGPFPDELNRNFERFYGTHNKIGNLPTYLQVENFLAAQSNIDSGNLLTDD
FDINLSWTRNNNSFFTPAGNHORAFYKMTVPGSDAQVFKLOYDVROYFPLTKHEF
TILLRLGRLGSGYQDCKDNLPFVENFYAGGFTSLRFGNSNAGKAVRYDSGS
NNGSDPATDSGGNAILASVELIYPTPFASEEARNQIRTSIFIDYDMASVWDEFDYR
GKADYGNQIYYDSDPTNRSSTGVALQWVSPMGPLVFLSLAKPIPKIKYDGDDEEFFFTT
IGRTF"
/complement(4118. .5476)
/gene="VC2253"
/complement(4118. .5476)
/gene="VC2253"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
```

```
/protein_id="AAF95397.1"
/db_xref="GI:9656814"
/translation="MDLILWNFAITIALGILVAVHEFGHFWARRCGVKVEPSIOF
GKSIWKRVGHDGTEYSIMPLGGYVKMLDGRVDDVPAEQQAADFQSLWKRSAIYS
AGPIFNFLFAIFAYWLVTGPAVKPIGVEVTPYSTAAQAGLEPGMEIKAVSGVNTP
DWESVNMGLIGHICDDSMITIVSSAEGVGLNEIKTILRDWNEFDPEESAGALGFRP
FTPEISNQLTNVSAQAGERAGLVQGTVLQINGQAVQVYVNAIQSHPNAPIAVM
VERAQOQVELTLIPDSRELQSGVQFAGIAPKVAEMPONVREPLGFGVPESLGKAVE
KSGQVLDLTVSMLKLLVGDVGLNNLSGPISIAGACTTADYGFVYPLGLALISINL
GIINLPLPMLDGGHLLFFMEAVIRRPVEKVOEMGIRIGGAILFSLMAVAIFNDET
RL"
/complement(5473. .6681)
/gene="VC2254"
/complement(5473. .6681)
/gene="VC2254"
/notes="similar to SP:P45568 GB:U00096 PID:1552750
PID:1786369; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="1-deoxy-D-xylulose 5-phosphate reductoisomerase"
/protein_id="AAF95398.1"
/db_xref="GI:9656815"
/translation="MRKLTILGATSGTSLKVIAQNPOQFSIVALVAGVNVAKMVO
LCOQWRPKYAVMATASAESEQLGLKNQAMATEVLYGEEAMCQVAALDDVDVTNMAIV
GAAGLLPTMAAVKAGKRVLLANKEALVMSQLFIDAVAQSGAELMPDSEHNAIFQCL
PTRIOTQGRCDLSOHGIDHILLTGSGGPFYSDLATLDSVTPEQATAHPNWSMGPKI
SVDSATMMNKGLLEYIEAKWLFNTSRLOLKVLIHQSVIHSVMQVQDCSVTAQLGEPDM
ATPISYAMAYPERVTAGVPALDFTRLQQLTFMEVDFAFYPCQLAMDACFLGQHATTS
LNAANEVAVDAFLKRIKRIFTDIALINDQVLSKVCATNTQLHCRDLESLLLELDTMARHF
AHQVLERQA"
/complement(6730. .7572)
/gene="VC2255"
/complement(6730. .7572)
/gene="VC2255"
/notes="similar to GB:L42023 SP:P44937 PID:1006039
PID:1221024 PID:1205166; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="phosphatidate cytidylyltransferase"
/protein_id="AAF95399.1"
/db_xref="GI:9656816"
/translation="MKQRIITALILAPVLILGILYLPFAFMALAVVTLILGFWEWTQ
FVNQSRMLAMIPALLVGGISVALIDFQFPAISNMNTAHFVLIGIGISLWLVSSGLAI
TYDRSPRLPHESSTVRHLGLFTLLPFNSVLELRADTYLSLDPYLGAKLVFLVCFLVW
AADSQAFYVGKSLGCKHMAPSPFNKTIEGLGVITAMLVGIWVWABCFCGQFQSPMPV
MLLIILLTVVISVLGDLVESMFRKVSIGIKDSSNIIPGHGILLDRIDSLSATAFPV FALL
YFLF"
/complement(7590. .8360)
/gene="VC2256"
/complement(7590. .8360)
/gene="VC2256"
/notes="similar to GB:L42023 SP:P44938 PID:1006041
PID:1221025 PID:1205167; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="undecaprenyl diphosphate synthase"
/protein_id="AAF95400.1"
/db_xref="GI:9656817"
/translation="MEVFLMQLSLSDQVLPKHIADVIMPNGRWAKAOCKPRVFGHK
NGVAAVRKTISSARLGIKAVTLPAESSENRRRPEEVGLMELPHTVLTSTEIKLHK
NULKRVIGDKSRFSERLQKHTMAQLPDVLDLIRITSGEGRISNFMQLWQLAYAEYFTP
KVTQGBELQSDIDEVFKQHLTMADLPDVLIRITSGEGRISNFMQLWQLAYAEYFTP
VWPFEDENCLIEAVTFWNRERRRFGCTGQVKALMEN"
/complement(8434. .8991)
/gene="VC2257"
/complement(8434. .8991)
/gene="VC2257"
/notes="similar to GB:L42023 SP:P44307 PID:1005819
PID:1220904 PID:1205056; identified by sequence
similarity; putative"
/codon_start=1
```

```

/transl_table=11
/product="ribosome recycling factor"
/protein_id="AAF95401.1"
/db_xref="GI:9656818"
/transl_table="MINEIKKDAQERMEKSVKALNGLSKIRTGRAHPSLLTGLISVDY
YGATPLNOVANVAEDARTLAIVFDRELQKVEKALMSDLGNPSAGTIIIRVPL
PPLTERRDIVKVRGAEGGRVAVNRIRNDANLKKLKDKEISDDERRAEQEI
QKLTDAAVKKIDDLVLAKEKELMEV"
/gene="VC2258"
/complement(9045. .9776)
/gene="VC2258"
/complement(9045. .9776)
/gene="VC2258"
/notes="similar to GB:D26562 SP:P29464 GB:D13334 GB:X78809
PID:1208944; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="uridylylate kinase"
/protein_id="AAF95402.1"
/db_xref="GI:9656819"

alignment_scores:
  Quality: 10.00 Length: 10
  Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AE004297/rev ..
Align seg 1/1 to reverse of: AE004297 from: 1 to: 10901

368 AspGluValLeuArgArgGluMetArgGln 377
|||||
2986 GATGAAGTCTTGGCGCGTGAATGCGTCAG 2957

seq_name: gb_ba:AE003941

seq_documentation_block:
LOCUS AE003941 11711 bp DNA linear BCT 15-JUN-2001
DEFINITION Xylella fastidiosa 9a5c, section 87 of 229 of the complete genome.
ACCESSION AE003941 AE003849
VERSION AE003941.1 GI:9105978
KEYWORDS
SOURCE
ORGANISM
Xylella fastidiosa 9a5c.
Xylella fastidiosa 9a5c
Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
Xylella.
1 (bases 1 to 11711)
Simpson,A.J., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
Alvarenga,R., Alves,L.M., Araya,J.E., Baia,G.S., Baptista,C.S.,
Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.,
Bueno,M.R., Camargo,A.A., Camargo,L.E., Carraro,D.M., Carrier,H.,
Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C., Costa-Neto,C.M.,
Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorry,H.,
Facincani,A.P., Ferreira,A.J., Ferreira,V.C., Ferro,J.A.,
Fraga,J.S., Franca,S.C., Franco,M.C., Frohme,M., Furlan,L.R.,
Garnier,M., Goldman,G.H., Goldman,M.H., Gomes,S.L., Gruber,A.,
Ho,P.L., Hohnselt,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P.
and Marino,C.L.
The genome sequence of the plant pathogen Xylella fastidiosa. The
Xylella fastidiosa Consortium of the Organization for Nucleotide
Sequencing and Analysis
Nature 406 (6792), 151-157 (2000)
JOURNAL MEDLINE PUBMED
20365717
10910347
REFERENCE
2 (bases 1 to 11711)
Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
Alvarenga,R., Alves,L.M.C., Araya,J.E., Baia,G.S., Baptista,C.S.,
Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.,
Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Carraro,D.M.,
Carrier,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C.,
Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E.,
Docena,C., El-Dorry,H., Facincani,A.P., Cristofani,M., Dias-Neto,E.,
Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C.,

```

```

Frohme,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,M.H.S.,
Gomes,S.L., Gruber,A., Ho,P.L., Hohnselt,J.D., Junqueira,M.L.,
Kemper,E.L., Kitajima,J.P., Krieger,J.E., Kuramae,E.E., Laigret,F.,
Lambais,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A.,
Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N.,
Madeira,H.M.F., Marino,C.L., Marques,M.V., Martins,E.A.L.,
Martins,E.M.F., Matsukuma,A.Y., Menck,C.F.M., Miracca,E.C.,
Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A.,
Nascimento,A.L.T.O., Netto,L.E.S., Nhani Jr., A., Nobrega,F.G.,
Nunes,L.R., Oliveira,M.A. de Oliveira,M.C., de Oliveira,R.C.,
Palmieri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira
Jr.,H.A., Pesquero,J.B., Quaggio,R.B., Roberto,P.G., Rodrigues,V.,
de M. Rosa,A.J., de Rosa Jr.,V.E., de Sa,R.G., Santelli,R.V.,
Sawasaki,H.E., da Silva,A.C.R., da Silva,F.R., da Silva,A.M., Silva
Jr.,W.A., da Silveira,J.F., Silvestri,M.L.Z., Siqueira,W.J., de
Souza,A.A., de Souza,A.P., Terenzi,M.F., Truffi,D., Tsai,S.M.,
Tsubako,M.H., Vallada,H., Van Sluys,M.A., Verjovski-Almeida,S.,
Vettore,A.L., Zago,M.A., Zatz,M., Meidanis,J. and Setubal,J.C.
Direct Submission
Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and
Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP
13083-970, Brazil
FEATURES
    source
        1..11711
            /organism="Xylella fastidiosa 9a5c"
            /db_xref="taxon:160492"
            /clone="9a5c"
            complement(84. .1424)
            /gene="XF1037"
            complement(84. .1424)
            /gene="XF1037"
            /note="similar to SPIP28183 (percent identity: 65 %/query
            alignment coverage: 100.0 %/subject alignment coverage:
            96.3 %); identified by sequence similarity; putative; ORF
            located using Glimmer/RBSfinder"
            /codon_start=1
            /transl_table=11
            /product="adenosylhomocysteinase"
            /protein_id="AAF83847.1"
            /db_xref="GI:9105979"
            /transl_table="MPCGLMSIRRVKASKQPLKGVRTGSLHMTIQTAVLIETLKDIGA
            DWRWASCNIFSTQDHAARAIATSGTPFAWKGTELEWYWDCTLAQLFTLSDGLTGP
            ELIVDDGDAFLTHKLENGSTLWDEQVIRKRLRIATERGEGYTRRVV
            NDWKGVEETTGHVRLYQIAATGRLLVPAINVDSVTKRFDNLGCGESLADGLKR
            AMDVLAGKLVAVCGYGDVGSASHRAYGVIVTEIDPICALQAMEGFEVTTVE
            DTLQADIVYTTGNKDVIRIEHTMAMKQDVIVCNIGDFNEIQVDALNALAGVQKIN
            IKPOVDKILPNGNTLFLAEGRLVNLGCATGHPFSVMSNSFANOTLAQIDLWQKDV
            YEKVYRLPKKLDEEVARLHLEKIGVKLTTLTANQAAYLGISVEGPKPEHRY"
            2470..3981
            /gene="XF1038"
            2470..3981
            /gene="XF1038"
            /note="similar to GI|2808752 (percent identity: 23 %/query
            alignment coverage: 104.4 %/subject alignment coverage:
            98.7 %); identified by sequence similarity; putative; ORF
            located using Glimmer/RBSfinder"
            /codon_start=1
            /transl_table=11
            /product="peptide synthase"
            /protein_id="AAF83848.1"
            /db_xref="GI:9105980"
            /translation="MVHCVCLVYVSKRCSECIHVCHYRRVAGLVLPVQCQTVSWCLHGD
            DLDLCYLMSLVAFSPSSGVASLPMGVSRLRPLVHVRKPIITLGRFLAQVRGLAAQL
            PSGRYVYNLCEDRYRFLAFACALRGVSLPSRAPGVVAEYIAYAPDSYCLGDR
            LSLRLRYWHLPEIFFELDGAIPSLSDAVVAIGTSGTGVPTSNLKTWGLVSTSE
            QDWAFHSLMGSOVTVVATVPPOHMYGMEESVLLPLLAPVAVVAGRPFPEDVAQAL
            AEVPATCTTAVRTPVHLRVQSGRLFRFEVAGIVSATPLTHDLAVAEACGCELRV
            FGSPETCLVAVTRFVQWTRFDVGSGLRFEVAGIVSATPLTHDLAVAEACGCELRV
            RFEVLHQADLLETAGRSLGELNRLCLAIIPGVYDGLVQMPLEADQVGRIVAAVV
            APTLHEVHIQKVLTKIDPLFLPRLRKVDALPRNETGKLPREALLALADVC"
            4249..4413
            /gene="XF1039"
            4249..4413
            /gene="XF1039"
            4249..4413

```

```

/gene="XF1039"
/translation="hypothesized protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothesized protein"
/protein_id="AAF83849.1"
/db_xref="GI:9105981"
/translation="MLVNTSVCNVSWLVFTQHCLLGLVAEMRCLLAILARWILVF
SMGCKYRGIV"
4430. 4591
/gene="XF1040"
4430. 4591
/gene="XF1040"
/translation="hypothesized protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothesized protein"
/protein_id="AAF83850.1"
/db_xref="GI:9105982"
/translation="MAFTGCAGCGVCTFFNGCVAGVQLGRGLMVCVIEWSVLGRSGFY
HSVSSGWS"
complement(4780. 5484)
/gene="XF1041"
complement(4780. 5484)
/gene="XF1041"
/translation="hypothesized protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothesized protein"
/protein_id="AAF83851.1"
/db_xref="GI:9105983"
/translation="MVSQDFTQHLPSNTALVAGVDEAGRPLAGPVVAAVVFDPFSQ
PRINGLDSKQLSPACRERLYAHIVERALAYKVMIDSTQIDTFLNIYQATMLGMRLAV
EGVAHVAKASIDGNRLPKNLPCPAEALVGGDARDITMAASILAKVTRDRHWLHLE
QYPHYGDFRKHGCTGAHLAALAEHGFCLERHRSFAPVRKMLTLEAHAKQSTHQHNE
NSTPKAFAFNMLLERDD"
complement(5541. 6698)
/gene="XF1042"
complement(5541. 6698)
/gene="XF1042"
/translation="hypothesized protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="lipid A disaccharide synthase"
/protein_id="AAF83852.1"
/db_xref="GI:9105984"
/translation="MIQAPRIALIAEASGDHLAGLIQQLRLRLPTAFVIGIGDDMM
RSARCTWTFDTELVAGLTGVLRLPKIRRECKRALAWHPDVLGIDAPDENL
TVRFWPKQRIHVTVHVSPIWAKREKRAKICASVDVLCLEPMPPEPIYARVIGIDAR
FVGHPADEPIPTQDARTALGCLLSPVLAVLGRSHSEISQSGSTFLEAAGLS
EHLPGHADVPIANTOCKPLLAQLSRSLPVMHSHLDNSARTAMLAADVILVSGT
ATEAMLLKRPWVAYKAPLTVRYVTKLKLKINRFPALPNTILAGEDLAPETQKDCI
APALCAALLDFKHPQKVATLQNRYLQLHTQLRRNASTRAAEAITELLPOR"
complement(6775. 7566)
/gene="XF1043"
complement(6775. 7566)
/gene="XF1043"
/translation="hypothesized protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="UDP-N-acetylglucosamine acyltransferase"
/protein_id="AAF83853.1"

```

```

/db_xref="GI:9105985"
/translation="MNKHASLIHPTAVIAPSATLPDVQIGAFTLIGNDVOIDGTII
GSUCTIHGPRICGNRRFIQAAAGGSPQDKKFAGETELLIGNNTIREVTTNRGT
GGGGVTSIGNDNWILAYTHIAHDHVGHCVFSSNNASLAGHVTVGDMVIFSGSAH
QFCRIGYAFIAGMTLGVDPFPTLIGSDTLGRPSGLKRRNFTPERITAIKR
AYRTLYVAGLPALAEAKQQAQAEAKDNDNDIKELQFIETAQRPLLR"
complement(7588. 8070)
/gene="XF1044"
complement(7588. 8070)
/gene="XF1044"
/translation="hypothesized protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="(3r)-hydroxymyristoyl ACP dehydrase"
/protein_id="AAF83854.1"

```

alignment\_scores:

Quality:	10.00	Length:	10
Ratio:	1.000	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

alignment\_block:

US-09-701-711-2 x AE003941/rev ..

Align seg 1/1 to reverse of: AE003941 from: 1 to: 11711

368 AspGluValLeuArgArgGluMetArgGln 377

|||||

10531 GATGAAGTACTACGTCGTAATGCGTCAG 10502

seq\_name: gb\_ba:AE004784

seq\_documentation\_block:

LOCUS AE004784

DEFINITION Pseudomonas aeruginosa PA01, section 345 of 529 of the complete genome.

ACCESSION AE004784 AE004091

VERSION AE004784.1 GI:9949799

KEYWORDS

SOURCE

ORGANISM

Pseudomonas aeruginosa.

Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

Pseudomonas.

REFERENCE

1 (bases 1 to 12469)

Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warrenner, P., Hickey, M.J., Brinkman, F.S., Hufnagle, W.O., Kowalik, D.J., Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R., Smith, K., Spencer, D., Wong, G.K., Wu, Z. and Paulsen, I.T. Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen

Nature 406 (6799), 959-964 (2000)

2 (bases 1 to 12469)

Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warrenner, P., Hickey, M.J., Brinkman, F.S.L., Hufnagle, W.O., Kowalik, D.J., Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H., Wong, G.K.-S., Wu, Z., Paulsen, I.T., Reizer, J., Sater, M.H., Hancock, R.E.W., Lory, S. and Olson, M.V. Direct Submission

Submitted (16-MAY-2000) Department of Medicine and Genetics, University of Washington Genome Center, University Of Washington, Box 352145, Seattle, WA 98195, USA

Location/Qualifiers

1. .12469

/organism="Pseudomonas aeruginosa"

/strain="PA01"

/db\_xref="taxon:287"

```

gene
complement(106..3627)
/feature="dnaE"
/feature="PA3640"
complement(106..3627)
/feature="dnaE"
/codon_start=1
/transl_table=11
/product="DNA polymerase III, alpha chain"
/protein_id="AAG07028.1"
/db_xref="GI:9949800"
/translation="MTVSFVHLRLHTEFSLVDGLVRVKPLAKAVAGLGMFAVAVTDQS
NMCSLVKFKVTAMGAIKPIGADILWASREDGDLRLSLILAMNAKGYRLTELISR
GWSEGORNGEIIIERDWDKEAREGLIALSAKEGEIGHALLDGEAKAEALLREWEV
FPEREYVQVRSRNDWEHLHAVALASRCNAPLVATNDVRFIKQEDFAEHETRVCI
GGRTILDPRRTSDQOYLKSPAEMAEFLSDFLEALENTVEIARCNIEVQOLKGYF
LPDFFPGMGIDDLIRHASFEGLERLEVLKPDTPDYEARQVYVDRINLELDII
QMGFFGYLIVMDIFKWKANGVPVPGRGSGAGSLVAVLKITDLDPLAYLDFERF
LNPERSIMPFDVDFCMGRDRIYVADAYGRNAVSOIITFTGNAKAAVVRDVARVO
GKSYGLADRLSKMIPFVCGMTLDKAYEQEEMLRDLKSDDEAAEIEWEAKLLEGITRG
TKKHAGGVVIATKTLDRSPACDEGGGLVTQFDKDDVEAAGLVKFDPLGLRITLI
KWAMELINREQAKGLEPNVIDFIPLDKPTYSLLQKAEFTAVFOLESGRMKELIKKL
KPDCLIEDJIALVLPFGPQLSQMVDDFINKHGRAELSPHPDYQAGLEPLVKPTY
GLILYQEOVMQIAQVMAGYTLGGADMLRRAMKKPEEMAKORGFGIECKNMGIDAD
LAGNIFDLVEKFAGYGNKSHSAAYGLVSYOTAWLKTHTPPAPFAAVLTADMDNDKV
VTLIRECHMKLRIIVAPDVNNSEFRFTVDDGRIVYGLGAIKGVGEGPEATECRAE
GGPNTLFDRCRVDLKRINKRTLEALIRAGALDRVGHYHDELKAYQATVDLNRVL
LAAMEEAQAEQARTASHSDSHMDLFGGVFAPEADYVANHKKVKELTKERLKGED
TGLYLTGHPIDEXEYGVRRFARQRIVELPARDTQTVAGLIVNLVRMKNKKGDMKF
VTLDDRSRIEASLSEAFAPAAQSLLOTDALVVVEGVSQDDFSGGLRLRAKRVMSLE
EARTGLASRLMKLHADLLKGLRLWLGELFNHRHRCGSCPITLDYTSADAKALLQFES
WRVDPADLLIOALRQFGCRDNLVNR"
complement(3757..5172)
/feature="PA3641"
complement(3757..5172)
/feature="PA3641"
/codon_start=1
/transl_table=11
/product="probable amino acid permease"
/protein_id="AAG07029.1"
/db_xref="GI:9949801"
/translation="MOEFATQVVDLLNGLIWIKILWLLVGCGLYFTVRLGLIQFRHF
GHTFSVLKGSQSDSGISSFOALCTSLAARVGTGNVAGVAVAITLGGPGAVFVWMAI
ALVGMATCFVETLAOLFKIRDDKQPRGGPAYMEKGLGARWMLFSIFLIADF
VNSVQANTITGAMOGAFVPTWISIAVAVILLIIFGGIARSARFSELAVFMAVA
YLLAVAGIILNSELPGVLAIVKSAFGWHEARAGGIGAAIINGFKKGLFSNEAGMG
SAPNAAGATPPHPASQGVQVMAGVFIDTLLICTASAAIILLAGPGEIGMILVQV
ALTQVGDWGRYFLAVIILFFAFTSIVANYFYAENCLVLEHNLHAGLLIFRLIYIAM
VMFGALASLPVWNLADYSMGLMATTNLIATILLISNLAIKLAKDYNAQRKAGKLTED
ASQFPEVQOKLEPGIWDORRA"
complement(5389..5994)
/feature="rnhB"
/feature="PA3642"
complement(5389..5994)
/feature="rnhB"
/codon_start=1
/transl_table=11
/product="ribonuclease HII"
/protein_id="AAG07030.1"
/db_xref="GI:9949802"
/translation="MQLGDFNLVEDLVAGVDEVRGRLCGPVVTAAVILDPSPRILG
LNDKSLSEARREALFEIREKALAWCIARAEEIDRLNLTILHATMLAQRAVSELSV
LPRALIDNGRCPLKAVPCAPVVKGDSQVPAALAAESILAKVSRDREWVELDRVTPGY
MAGHKGYPTAVHLKALSRIGTPPIHRRSFAPVRELLDVSQ"
complement(5994..7130)
/feature="lpxB"
/feature="PA3643"
complement(5994..7130)
/feature="lpxB"
/codon_start=1
/transl_table=11
/product="lipid A-disaccharide synthase"
/protein_id="AAG07031.1"
/db_xref="GI:9949803"
/translation="MADGLRVALVAGEASGDILGSLMQALRARHPDIEFIGVGGPRM
EAEGLSSYPMERLTSYVGVLEVLGRPELRLRRKRLRIETLIEARPDMVTGIDAEFTL
GVEHKLROAGLRTVYVSPSWAWRQRLVKIAREADLMALFPFPAFYEEHGVPR
VGHPLANTIPLOADRAARARGLPADGOVLALMPGSGGEGVKGALGADFLDIAORLL
VERPGILRVLPCLASARREOIEQMLQREPLTLTLDDGASHEALACADVLIASTAT
LEALLYKRPMVAVYRAGLTYRILKRLKSPYISLPLNLAGLLVPELIQDAATPOAL
AATLSPLDDGSGQVEFFDAIHRALRQDASAAAEAVLQVLR"
complement(7134..7910)
/feature="lpxA"
/feature="PA3644"
complement(7134..7910)
/feature="lpxA"
/codon_start=1
/transl_table=11
/product="UDP-N-acetylglucosamine acyltransferase"
/protein_id="AAG07032.1"
/db_xref="GI:9949804"
/translation="MSLIDPRAIIDSARLAADVQVPMISVGAEEVEIGEGVIGPHV
VLKPTKIKGNRIYQFSSVEDTDLKYKGETRLVIGDHNVIREGTVIHRGTVOCDR
AETTTGDHNLIMAYAHIGDSVIGNHCILVNTLALAGHVDDWALISYTLVHOYCR
IGAHSFSGMSAIGKDVPAIVTVFVGNPAEASRMFEGMRGRGFSSEAIHALRRAYKV
YROGHTVEALAEALAESAAQFPEVAVFRDSIQSATRGITR"
complement(7907..8347)
/feature="fabZ"
/feature="PA3645"
complement(7907..8347)
/feature="fabZ"
/codon_start=1
/transl_table=11
/product="(3R)-hydroxymyristoyl-[acyl carrier protein]"
/protein_id="AAG07033.1"
/db_xref="GI:9949805"
/translation="MMDIENIREYLPHPYFLVDRVVELDIEGKRIRAYKNVSINEP
FPNGHPPHPIMPVGLIETAMAQAAGILGKMLDVKPADGTLYFYFGSKDLRFQPV
PGDQLQHLAKFTSVKRSIWKFDCHATVDDKPVCSAEIICAERKL"
complement(8393..9454)
/feature="lpxD"
/feature="PA3646"
complement(8393..9454)
/feature="lpxD"
/codon_start=1
/product="UDP-3-O-[3-hydroxylauroyl]"
/protein_id="AAG07034.1"
/db_xref="GI:9949806"
/translation="NMSTLSYTLGQLAAHVGAVERGDADLPQGLATLQEAQPAQLSF
LANPYRKYLPESRAGVALLTAADAGFAGTAVANPYLAVASLSHLPDRKPKAAAG
THPTAIVADAEDVPSAGVAVIESGARICAGVSGAHCVIGARSYDIEGEGWLAAPR
VTLYHDVTIGARVSIQSGAVIGGEGGFANEGKGVOKIAQIGGVITGDDVEIGANTTI
DRGALSDTLIGNVKLDNQIMIAHNQIGDHTAMAACVIGSGSAIGRHCMLAGGVGL
VGHIEICDNVFTGMTVTRSITEPGSYSSGTAMQPAAEWKKSAARIQRLDDMARRLQ
OLEKRLAAVTSSGDASSDA"
complement(9454..9960)
/feature="PA3647"
complement(9454..9960)
/feature="PA3647"
/codon_start=1
/transl_table=11
/product="probable outer membrane protein precursor"
/protein_id="AAG07035.1"
/db_xref="GI:9949807"
/translation="MRKTFQVLIITAAAPSAFAEMKIAVLYNOMALLESDAKQY
AVDAKFKGPQNLKLNLERDALQDLVNSGSKSQSDREKALDFKQKARDFQFQ
SKELNESKAAADRLMKLKPDKQAVEETIKKGYDMVIERGAVVDKPYDITRQV
IERMNLNR"
complement(10011..12404)
/feature="PA3648"
complement(10011..12404)
/feature="PA3648"
/codon_start=1
/transl_table=11
/product="probable outer membrane protein"
/protein_id="AAG07036.1"

```



```
/db_xref="GI:9949808"
/translation="MKRFLLPKLLSALMIAEVAHSAFTVSDIRVNGLQVRSAGSVEAA
LPNVGETDDAALVQATRLFKTGFQDIOGLRGNVLVTVVVERPSISSTIEIENGK
ATSKEDLLKGLKSGLAEGEIQFQATLEGRNELQRYVAQGRYSAEINAEVTPQPN
RVALKINNEGIVAIASHINVGNTVFSSEDLDFELKATLNLWFFKNDKDKTAREKL
KVALERLRYLDRGQVSRKVTMTSDILTRRLGNEGYTFANVNGVPEAHDDKRTV
KYPEEVKRLLVQKQGVSRKVTMTSDILTRRLGNEGYTFANVNGVPEAHDDKRTV
SVTFVDPGKRAYVNRINFRNGTKTEDEVLRRMOMEGGWASTYLDGSKARLERLG
YKENVNTPAYGTDQDVNVNVEQPSGTSATSVGAQAGLILGGSISONNPLG
TKNKSIGTFRSEYOTRYNFGVDPVYTVYDGVSLGYNAYRTDDELDVDAVSVN
SLGAGMSIGPISESRRLYGLSVQRDQIDTGRYTVDEIYDFDKKGDNTFKASIG
WSESTLNGKVLATRHQSQTLETLPLGSDLSFYKIDYRGQVFAPLTDNTYRHFTEL
GCGDGYTERLPFYENYVAGFNSVRGKSTLGRSPFSVARNPDPGPMKNQGPDS
KGRYTDPODPAFGNLLITGCAELLPLFPFKDQRLTVLFWDVGSTFTDTCPTK
TTTNCIGIKTDNLASSGVGLTWITALGPLSLFSLATPIKKPDNAETQVQFSLGQTF"
BASE COUNT      2144 a  3895 c  4089 g  2341 t
ORIGIN

alignment_scores:
  Quality:   10.00      Length:      10
  Ratio:     1.000      Gaps:         0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AE004784/rev ..
Align seg 1/1 to reverse of: AE004784 from: 1 to: 12469
368 AspGluValLeuArgArgGluMetArgGln 377
|||||
11324 GACGAAGTGTCTGCGCGGAAATGCGCAC 11295

seq_name: gb_ba:AE000127

seq_documentation_block:
LOCUS      AE000127      14877 bp      DNA      linear      BCT 01-DEC-2000
DEFINITION Escherichia coli K12 MG1655 section 17 of 400 of the complete
            genome.
ACCESSION  AE000127 U00096
VERSION    AE000127.1 GI:1786370
KEYWORDS
SOURCE     Escherichia coli K12.
            Escherichia coli K12
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
            1 (bases 1 to 14877)
            Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
            Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
            Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
            Mau, B. and Shao, Y.
            The complete genome sequence of Escherichia coli K-12
            Science 277 (5331), 1453-1474 (1997)
            97426617
            MEDLINE
            PUBLISHED
            9278503
REFERENCE  2 (bases 1 to 14877)
            Blattner, F.R.
            Direct Submission
            Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Hall, Madison, WI 53706, USA.
            Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
            3 (bases 1 to 14877)
            Blattner, F.R.
            Direct Submission
            Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Hall, Madison, WI 53706, USA.
            Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
            4 (bases 1 to 14877)
            Plunkett, G. III.
            Direct Submission
            Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Hall, Madison, WI 53706, USA
```

## COMMENT

This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using Genemark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (<http://cgsc.biology.yale.edu>). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (<http://www.genetics.wisc.edu>). \*\*\* The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

## FEATURES

source	Location/Qualifiers
	1. .14877
	/organism="Escherichia coli K12"
	/strain="K12"
	/sub_strain="MG1655"
	/db_xref="taxon:83333"
	123. .884
gene	/gene="yaeS"
	/note="b0174"
	123. .884
CDS	/gene="yaeS"
	/function="orf; Unknown"
	/note="0253; 61 pct identical to 229 residues of an approx.240 aa hypothetical protein Y920_HAEIN SW: P44938"
	/codon_start=1
	/transl_table=11
	/product="orf, hypothetical protein"
	/protein_id="AAC73285.1"
	/db_xref="GI:1786371"
	/translation="MMSATQPLSEKLPAGCHRVAILMDNGNRWAKKQKIRAFGKH AGAKSVRAVSFAANNQIEALTLYAFSSNNRPAQEVSAIMELFYWALDSEVKSLLHR HNVRLRIIGDTSFNSRLQRIKSEALTAGNTGLTLNTAANYGGRWIDVQGVQLAE KVOQGNLPQDQIDEMLNQHVCHELAPDLVIRGTGHEIRISNFWLLQIAYAEIYFTD VLWPDFDEQDFEGALNAFANRRRRFGTPEGDETA"
promoter	870. .900
	/note="factor Sigma70; predicted +1 start at 195687"
promoter	899. .927
	/note="factor Sigma70; predicted +1 start at 195714"
promoter	956. .987
	/note="factor Sigma70; predicted +1 start at 195774"
gene	1005. .1754
	/gene="cdsA"
	/note="b0175"
	1005. .1754
CDS	/gene="cdsA"
	/EC_number="2.7.7.41"
	/function="enzyme; Fatty acid and phosphatidic acid biosynthesis"
	/note="0249; 100 pct identical to CDS_A_ECOLI SW: P06466"
	/codon_start=1
	/transl_table=11
	/product="CDP-diglyceride synthetase"
	/protein_id="AAC73286.1"
	/db_xref="GI:1786372"
	/translation="MLAAWAEWQSGFTTSQRVWLAIVLGLLLALMLFLPEYHRNI HOPLVEISLWASLGWIVALLLVFPGSAAIWRNSKTLRILFGLVLTVPFFWGMAL

```
protein_bind
RAWHYDNHSGAIWLIVMLVWGADSGAYMGKLFQKHKHKLAPKVPSPGKTWQGFIGG
LATAAVISWGYGMANLDVAPVTLICSIIVAALASVLGDLTFESMFKREAGDKSGHLI
PGHGILDRIDSLTAAPVFPACLLLVFRTL"
1421..1432
/gene="cdsA"
/note="No predicted promoter"
/bound_motety="PutA predicted site"
1766..3118
/gene="yael"
/note="b0176"
1766..3118
/gene="yael"
/function="orf; Unknown"
/note="o450; 100 pct identical to fragment YAEI_ECOLI
SW:P37764 but contains 302 additional C-terminal residues"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC73287.1"
/db_xref="GI:1786373"
/transl_table="MLGFLDLASFIVALGVLTIVHEFGHFVARRCGVRVERFSIGF
GKALRRDCKLGTGVVIALPLGCVKMLDERAEPVPELRHHAFNNKSVGORAAIIA
AGPVANIFALFAWLVFIIGVPGVPRVWGEIANSIAAEQAIPAGIELKAVDGIETP
DWDVRLQVLDKIDESTITIVAPFSDQRDKVLDLRHNAFEPDKEDPVSLSGIRPR
GPOTEPLENVQPNASAKAGLQAGDRIVKDGQPLAQWTFVLMVLRNPGKSLALEI
ERQSPSLSTLIPESKGAIGFVIEPKVILPDEYKVVROGYPFNIAVEATDKT
WOLMKLTVSMGLKITGDVKNLNLGSPISIAKGAGTAELGVVYLPFLALISVNLGI
INLPLPLVDGSHLLFLAIEKIKGPPVSERVQDFCYRIGSGILLVLLGLALFNDFSRL
"
3148..5580
/gene="yael"
/note="b0177"
3148..5580
/gene="yael"
/function="orf; Unknown"
/note="o810; 45 pct identical (29 gaps) to 808 residues of
an approx. 800 aa protein D151_HAEIN SW: P46024"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC73288.1"
/db_xref="GI:1786374"
/transl_table="MAMKLLIASLFSSTVYGAEGFVVKDIHFEGLRVAVGAALL
SMPTVGTDTNDEDSITRALFATGNFEDVRVLRDGDITLIVQKERTIASIFSGN
KSVKDDMLKNLEASGVVRGESLDRITIADEKLEDFYXSVGKYSASVAVVPLPR
NRVLKLVFQGVGSAEQIINIVGNHAFITDELISHFQLRDEVPWNVGDRKYOKK
LAGLETLRSYLRDGRARENIDSTQVSLPDKGIYVYNTIEDQYKLSGVYVSGN
LAGHSAIEQLTKPELGELNGTKVTRMEDDIKLLGRYGTAYPRVSPMEINDADKT
VKLRVNDAGNRFTYVRKIRFEGNDSKDAVLRREMRQMEGAWGLSDLDQGERLNL
GFFETVDTQRPVGPQDVVYKVKERNTGSPFNFGIGYTESGVSFQAGVQKQNL
GTGVANGTNGKNDQTYAELSVNTPYFVDVGSGLGRFYNDFOADDADLSQDYNKS
YGTDTVLGFFINENYSLRAGLVYVHNSLNMPQVAMRVLYLSMGEHPSTSDQNSFK
TDDTFYNGWTYKLDKRGYPTDGRVNLGKVTIPGSDNEYKYKVLDTATVYPIDDD
HKWVYLRTRWGYDGLGKGMPEYFNFYAGGSTVRFQSGNTIGKRAVTFPHQASNY
DPDYDECATODGAKDLCKSDDAVGNAMAVASLEFITPTFFISDKYANSVRSFFWD
MGTVWDMFNWDSQYSGYPDYSNIRMSAGIALQWMSPLGLVFSYAQPFKKYDGGK
AEQFQFNIGKTW"
5583..5611
/note="factor Sigma70; predicted +1 start at 200398"
5702..6187
/gene="hlpA"
/note="b0178"
5702..6187
/gene="hlpA"
/function="factor; Basic proteins - synthesis,
modification"
/note="o616; 100 pct identical to HLP_A_ECOLI SW:
P11457; similar to ompH of other bacterial species"
/codon_start=1
/transl_table=11
/product="histone-like protein, located in outer membrane
or nucleoid"
/protein_id="AAC73289.1"
```

```
/db_xref="GI:1786375"
/transl_table="MKKWLALQAGLGLATLQAQADKIAIVNMGSLFQOQVAKQKTGVSN
TLENEFKGRASELQRMETDLOAKMKKQSMKAGSDRTKLEKDVMAOROTFAQKAAQAF
QDRARRSNEERGKLVTRIQATVAKSVANSQDIDLVDANAVAYNSSDVKDITADVLKQV
K"
6191..7216
/gene="lpxD"
/note="b0179"
6191..7216
/gene="lpxD"
/EC_number="2.3.1.-"
/function="enzyme; Surface polysaccharides and antigens"
/note="o341; 100 pct identical to LPXD_ECOLI SW: P21645
but includes initiator met"
/codon_start=1
/transl_table=11
alignment_scores
Quality: 10.00 Length: 10
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block
US-09-701-711-2 x AE000127 ..
Align seg 1/1 to: AE000127 from: 1 to: 14877
648 PropHetYrGluAsnPhetYrAlaGlyGly 657
|||||
5086 CGGTCTACGAGAACTTCTATGCCGTGGT 5115
seq_name: gb_ba:AE008705
seq_documentation_block:
LOCUS AE008705 22997 bp DNA linear BCT 25-OCT-2001
DEFINITION Salmonella typhimurium LT2, section 13 of 224 of the complete
genome.
ACCESSION AE008705 AE006468
VERSION AE008705.1 GI:16418723
KEYWORDS
SOURCE
Salmonella typhimurium LT2.
Salmonella typhimurium LT2
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Salmonella.
REFERENCE
1 (bases 1 to 22997)
McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,
Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Flores,L., Miller,W.,
Stoneking,T., Nhan,M., Waterston,R. and Wilson,R.K.
Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2
JOURNAL
NATURE 413 (6858), 852-856 (2001)
PUBMED
11677609
REFERENCE
2 (bases 1 to 22997)
The Salmonella typhimurium Genome Sequencing Project.
Direct Submission
Submitted (29-MAR-2001) Genome Sequencing Center, Department of
Genetics, Washington University School of Medicine, 4444 Forest
Park Boulevard, St. Louis, MO 63108, USA
Supported by NIH grant 5U 01 AI43283
COMMENT
Coding sequences below are predicted from manually evaluated
computer analysis, using similarity information and the programs;
GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and
GeneMark; http://opal.biology.gatech.edu/genemark/
```

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>, and Pedro Romero and Peter Karp at EcoCyc; <http://ecocyc.PangeaSystems.com/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites

were kindly provided by Heladia Salgado, Julio Collado-Vides and ReguonDB;  
http://kinich.cifn.unam.mx:8850/db/regulondb\_intro.frameset

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistries or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one ml3 subclone.

FEATURES	source
	location/Qualifiers
	1. .22997
	/organism="Salmonella typhimurium LT2"
	/strain="LT2; SGSC 1412; ATCC 700720"
	/db_xref="ATCC:700720"
	/db_xref="taxon:99287"
	98. .103
	/gene="frr"
	/note="Putative RBS for frr; RegulonDB:STMS1H000586"
	98. .668
	/gene="frr"
	/note="STM0219"
	111. .668
	/gene="frr"
	/note="Ortholog of E. coli ribosome releasing factor (AAC73283.1); Blastp hit to AAC73283.1 (185 aa), 93% identity in aa 1 - 185"
	/codon_start=1
	/transl_table=11
	/product="ribosome releasing factor"
	/protein_id="AAL19183.1"
	/db_xref="GI:16418724"
	/translation="MISDIRDAEVRMEKVEAFKTIQISKVRTGRASPSLDGIWVEY YGPTPLQRLASVTVEDSRTLKINVEDRSMPAVEKAIMASDLGLNPSSAGTDIRPL PPLTEERKDLTIKIVRGEAEQARVAVNRVRDANKVKALLKKAISEDDDRSEQEV QKMTDAIKKVAALADKAEALMQF"
	798. .2005
	/gene="dxr"
	/note="STM0220"
	798. .803
	/gene="dxr"
	/note="Putative RBS for dxr; RegulonDB:STMS1H000587"
	809. .2005
	/gene="dxr"
	/EC_number="1.1.1.1"
	/note="Ortholog of E. coli putative ATP-binding component of a transport system (AAC73284.1); Blastp hit to AAC73284.1 (398 aa), 88% identity in aa 1 - 397"
	/codon_start=1
	/transl_table=11
	/product="1-deoxy-D-xylulose 5-phosphate reductoisomerase"
	/protein_id="AAL19184.1"
	/db_xref="GI:16418725"
	/translation="MKQLTILGSTGTCSTLDVVRHNPDSFRVIALVCKNVMARAE QCLFSPRYAVMDTSSAEQLKIMLQHGSRTEVLSGQAAACEAALDEGHVMAAIV GAAGLPLTAAIRAGKTIILANKESLVTCGRLEFMEDEVKRSNARLLPVDSPHNAIFPSL POSIOHNLVADLEQCVTSILLTSGGPERETPCMDLAAMTPDOACRHPNWSMGRIK SVSATAHMKNGLEYIARWLFNASARQMEVLIHPOSVIHSWVYQGSVLAOLGEPDM RTPIATMAMPNVRTSGAQLDFCKLSALTFSAFDYQRYEPLCLKLANEAFQGGQAATA LNAAEITVAALFQAQQIRFTDIAGLNLAVLERMDLQEPASVEDVILQVDAIAREVARKQ VIRLSR"
	2299. .2304
	/gene="upps"
	/note="Putative RBS for upps; RegulonDB:STMS1H000588"
	2299. .3076
	/gene="upps"
	/note="STM0221"
	2318. .3076
	/gene="upps"
	/EC_number="2.5.1.31"
	/note="Ortholog of E. coli orf, hypothetical protein (AAC73285.1); Blastp hit to AAC73285.1 (253 aa), 93% identity in aa 2 - 253"

	/codon_start=1
	/transl_table=11
	/product="undecaprenyl pyrophosphate synthetase (dl-trans,poly-cis-decaprenylcistransferase)"
	/protein_id="AAL19185.1"
	/db_xref="GI:16418726"
	/translation="MLSATOPVSENLPAGRCRHVAIIMDGNRWAKKQGIKRAFQHKH GAKSVRRVAFSAANGIDALTLYAFSENWNRPAQEYSALMELFWALDSVKSILHRH NVRLRIIGDISRFNSRLOERIKSEALTHTNTGLTATANYGGRWDIVQVROLAEQ VQAGVLAPDQIDEERLGGQICMHELAPVDLVITGGEHRISNFLWQIAYAEILFTDV LMPDFDEQFEGLAHAFANRRERFEGGTEPGDDKA"
	3076. .3081
	/gene="cdsA"
	/note="Putative RBS for cdsA; RegulonDB:STMS1H000589"
	3076. .3946
	/gene="cdsA"
	/note="STM0222"
	3089. .3946
	/gene="cdsA"
	/EC_number="2.7.7.41"
	/note="Ortholog of E. coli CDP-diglyceride synthetase (AAC73286.1); Blastp hit to AAC73286.1 (249 aa), 95% identity in aa 1 - 249"
	/codon_start=1
	/transl_table=11
	/product="CDP-diglyceride synthase"
	/protein_id="AAL19186.1"
	/db_xref="GI:16418727"
	/translation="MLKYLGISAFVLIPAVTAAFLPLPPVGFATITLVVCMALAWENG QLSGFAARSQRVWLAVLCGLLALLFLLEPHHNIROPVEMSLWASLGMWVALLL VLFPYGSAAITWRNSKTLRIPLGLTIVPFFWGLALRAWHYDHSYGAIWLLVYML VWGADGAYMFGKLFKHKLAPVSPGKTQGFIGGLATAAIVISWGYGMANLNVAPV ILLICSVVAALASVLGDLTSMFKEAGINDSGHLIPCHGGLDRIDSLTAAPVFAF LILLVFPFL"
	3950. .5310
	/gene="yael"
	/note="STM0223"
	3950. .3955
	/gene="yael"
	/note="Putative RBS for yael; RegulonDB:STMS1H000590"
	3958. .5310
	/gene="yael"
	/note="Ortholog of E. coli orf, hypothetical protein (AAC73287.1); Blastp hit to AAC73287.1 (450 aa), 88% identity in aa 1 - 450"
	/codon_start=1
	/transl_table=11
	/product="putative membrane-associated Zn-dependent protease"
	/protein_id="AAL19187.1"
	/db_xref="GI:16418728"
	/translation="MLSILMNLAFFIATLGVLTIVHEFGHEFWARRCQVVERFSIGF GKALWRTDRYGTGYEVIALIPGGVYKMLDERAEPVAPELRRHAFNNKTVQORAAIIA AGPVANFIFAIFAYVWVFIIGVGVVRVIGETIPNSIAAQAIAPGTAKVADGIETP DWAIVRLQVSKIGDQTTVSVAQPSDQKQDQTLDLRHNWAFEPKQDPVSSLGIRPR GPQTPVRLSVSQANSAASKAGLQAGDRIVKVDGQPLQMMKFVTFVRDNPCKGLALEI BROGSLALTLPTDKTSVNGKAGFAGVWPKIIPLEPREYKTIROYGPFSAILEATDKT WQMLKFSVLMGLKITGDVLLNNLSGPISTAOAGMSAEFGVIVYLMFLALISVNLGI INLPLPVLGGHLLFLAIEKLGKGPVSERVDQFSYRIGSLILLVLLMGLALFNDFSLR "
	5342. .7756
	/gene="yael"
	/note="STM0224"
	5342. .7756
	/gene="yael"
	/note="Ortholog of E. coli orf, hypothetical protein (AAC73288.1); Blastp hit to AAC73288.1 (810 aa), 92% identity in aa 1 - 810"
	/codon_start=1
	/transl_table=11
	/product="putative outer membrane antigen"
	/protein_id="AAL19188.1"
	/db_xref="GI:16418729"

/translation="MAWKLLIASLLFSSATVYGAEGFVVVKDHFEGLRQVAVGAALL  
SPMVGDTVDNEDISITRALFATGNFEDRVLRDGNLTLLVQVKERTIASITFSGN  
KSVKDDMLKQNLASGVYGESLDRFTLSIEKGLDFYLSVGKYSASVAVVTPUPR  
NRVDLKLVEQVSAKIQOINIVGNHAFSTEELISHFOLRDEVPWNVVGDYKQOK  
LAGDLTLRSYLLDRGYARFNIDSTQVSLTPDKKGIYIVNTEGDOYKLSGVQVSGN  
LAGHSAEINLKIIEPGELYNGTKVTKMEDDIKKLLRGYIAYPRVQSQPEINDADKT  
VLRNVNAGNRVYRKIRFENDYTKSVLRREMRMEGAWLGSGLVDOGRKLNRL  
GFFETVDTDRVPSQDVVYKVSNTKSTFNFNGIGYTESGVFQGVQDQNL  
GTCYSVINGTNDQYSELVNTFYFVDMGSLGRIFYNDFOADDADLSDYTNKS  
YGDVTLFPINENYTLRAGLVVHKNLSVMOQJLAMDYVLESMSQSDATSSFAADDF  
TENYGWYKLDGRVPEPTDGSRVNLTKGVTFICSDNEYKYSILDTATYVPIDNHKKW  
VLGRHWGDLGLGKEMFFIENYTAGSSSTVRGOSNTIGPKAVYKNGAHTSWDDND  
DYEDCTQESGCKSDDAVGNMAVASLEFITTPTISEKYANSVRTSFFWMDGTWMDT  
NNDPSAPSDVPDYSDPGNIRMSAGIALQWMSPLGLVFSYAQPFKKYDCDRAEFOF  
NICKTW"

alignment\_scores:  
Quality: 10.00 Length: 10  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-701-711-2 x AE008705

Align seg 1/1 to: AE008705 from: 1 to: 22997

648 ProPheTyrGluAsnPhETyRAlaGlyCly 657

|||||

7268 CCGTCTCATGAGAACTCTATGCCGGTGGT 7297

seq\_name: gb\_htg:AC104418

seq\_documentation\_block:

LOCUS AC104418 62088 bp DNA linear HTG 10-DEC-2001  
DEFINITION Homo sapiens chromosome 15 clone RP11-1113M11 map 15, LOW-PASS  
SEQUENCE SAMPLING.

ACCESSION AC104418

VERSION AC104418.1 GI:17432864

KEYWORDS HTG; HTGS\_PHASE0.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 62088)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens chromosome 15, clone RP11-1113M11

Unpublished

REFERENCE 2 (bases 1 to 62088)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barina,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,  
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,  
Choepe,I., Colangelo,M., Collins,S., Collins,S., Cook,A.,  
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,  
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Heaford,A., Horton,L., Hulme,W., Illiev,I., Johnson,R.,  
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,  
Lamarez,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,  
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,  
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,  
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,  
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Trigliio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (10-DEC-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L21692  
Center clone name: 1113\_M\_11  
-----

\* NOTE: This record contains 79 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 677: contig of 677 bp in length  
678 777: gap of 100 bp  
778 1456: contig of 679 bp in length  
1457 1556: gap of 100 bp  
1557 2244: contig of 688 bp in length  
2245 2344: gap of 100 bp  
2345 3033: contig of 689 bp in length  
3034 3133: gap of 100 bp  
3134 3843: contig of 710 bp in length  
3844 3943: gap of 100 bp  
3944 4634: contig of 691 bp in length  
4635 4734: gap of 100 bp  
4735 5383: contig of 649 bp in length  
5384 5483: gap of 100 bp  
5484 6148: contig of 665 bp in length  
6149 6248: gap of 100 bp  
6249 6928: contig of 680 bp in length  
6929 7028: gap of 100 bp  
7029 7734: contig of 706 bp in length  
7735 7834: gap of 100 bp  
7835 8519: contig of 685 bp in length  
8520 8619: gap of 100 bp  
8620 9294: contig of 675 bp in length  
9295 9394: gap of 100 bp  
9395 10093: contig of 699 bp in length  
10094 10193: gap of 100 bp  
10194 10875: contig of 682 bp in length  
10876 10975: gap of 100 bp  
10976 11675: contig of 700 bp in length  
11676 11775: gap of 100 bp  
11776 12448: contig of 673 bp in length  
12449 12548: gap of 100 bp  
12549 13247: contig of 699 bp in length  
13248 13347: gap of 100 bp  
13348 14027: contig of 680 bp in length  
14028 14127: gap of 100 bp  
14128 14828: contig of 701 bp in length  
14829 14928: gap of 100 bp  
14929 15623: contig of 695 bp in length  
15624 15723: gap of 100 bp  
15724 16413: contig of 690 bp in length  
16414 16513: gap of 100 bp  
16514 17205: contig of 692 bp in length  
17206 17305: gap of 100 bp  
17306 17994: contig of 689 bp in length  
17995 18094: gap of 100 bp  
18095 18766: contig of 672 bp in length  
18767 18866: gap of 100 bp  
18867 19576: contig of 710 bp in length  
19577 19676: gap of 100 bp

```
* 19677 20364: contig of 688 bp in length
* 20365 20464: gap of 100 bp
* 20465 21162: contig of 698 bp in length
* 21163 21262: gap of 100 bp
* 21263 21953: contig of 690 bp in length
* 21953 22052: gap of 100 bp
* 22053 22740: contig of 688 bp in length
* 22741 22840: gap of 100 bp
* 22841 23539: contig of 699 bp in length
* 23540 23639: gap of 100 bp
* 23640 24346: contig of 707 bp in length
* 24347 24446: gap of 100 bp
* 24447 25138: contig of 692 bp in length
* 25139 25238: gap of 100 bp
* 25239 25909: contig of 671 bp in length
* 25910 26009: gap of 100 bp
* 26010 26694: contig of 685 bp in length
* 26695 26794: gap of 100 bp
* 26795 27482: contig of 688 bp in length
* 27483 27582: gap of 100 bp
* 27583 28279: contig of 697 bp in length
* 28280 28379: gap of 100 bp
* 28380 29076: contig of 697 bp in length
* 29077 29176: gap of 100 bp
* 29177 29846: contig of 670 bp in length
* 29847 29946: gap of 100 bp
* 29947 30618: contig of 672 bp in length
* 30619 30718: gap of 100 bp
* 30719 31410: contig of 692 bp in length
* 31411 31510: gap of 100 bp
* 31511 32184: contig of 674 bp in length
* 32185 32284: gap of 100 bp
* 32285 32961: contig of 677 bp in length
* 32962 33061: gap of 100 bp
* 33062 33738: contig of 677 bp in length
* 33739 33838: gap of 100 bp
* 33839 34545: contig of 707 bp in length
* 34546 34645: gap of 100 bp
* 34646 35343: contig of 698 bp in length
* 35344 35443: gap of 100 bp
* 35444 36096: contig of 653 bp in length
* 36097 36196: gap of 100 bp
* 36197 36892: contig of 696 bp in length
* 36893 36992: gap of 100 bp
* 36993 37678: contig of 686 bp in length
* 37679 37778: gap of 100 bp
* 37779 38459: contig of 681 bp in length
* 38460 38559: gap of 100 bp
* 38560 39264: contig of 705 bp in length
* 39265 39364: gap of 100 bp
* 39365 40038: contig of 674 bp in length
* 40039 40138: gap of 100 bp
* 40139 40809: contig of 671 bp in length
* 40810 40909: gap of 100 bp
* 40910 41593: contig of 684 bp in length
* 41594 41693: gap of 100 bp
* 41694 42373: contig of 680 bp in length
* 42374 42473: gap of 100 bp
* 42474 43168: contig of 695 bp in length
* 43169 43268: gap of 100 bp
* 43269 43966: contig of 698 bp in length
* 43967 44066: gap of 100 bp
* 44067 44761: contig of 695 bp in length
* 44762 44861: gap of 100 bp
* 44862 45541: contig of 680 bp in length
* 45542 45641: gap of 100 bp
* 45642 46349: contig of 708 bp in length
* 46350 46449: gap of 100 bp
* 46450 47138: contig of 689 bp in length
* 47139 47238: gap of 100 bp
* 47239 47939: contig of 701 bp in length
* 47940 48039: gap of 100 bp
* 48040 48697: contig of 658 bp in length
```

```
* 48698 48797: gap of 100 bp
* 48798 49471: contig of 674 bp in length
* 49472 49571: gap of 100 bp
* 49572 50253: contig of 684 bp in length
* 50256 50355: gap of 100 bp
* 50356 51055: contig of 700 bp in length
* 51056 51155: gap of 100 bp
* 51156 51844: contig of 689 bp in length
* 51845 51944: gap of 100 bp
* 51945 52626: contig of 682 bp in length
* 52627 52726: gap of 100 bp
* 52727 53401: contig of 675 bp in length
* 53402 53501: gap of 100 bp
* 53502 54198: contig of 697 bp in length

alignment_scores:
  Quality: 10.00      Length: 10
  Ratio: 1.000       Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AC104418 ..
Align seg 1/1 to: AC104418 from: 1 to: 62088
430 GlySerSerThrIleAlaAlaGlyTyrSer 439
|||||
21572 GCGAGCTCCACTATAGCAGCGGGTTACAGC 21601

seq_name: gb_htg:AC017607

seq_documentation_block:
LOCUS AC017607 81255 bp DNA linear HTG 10-DEC-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC017607
VERSION AC017607.1 GI:6554390
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 81255)
AUTHORS Adams, M. and Venter, J.C.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10211447 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
  source
    1..81255
    /organism="Drosophila melanogaster"
    /db_xref="taxon:7227"
BASE COUNT 22419 a 18447 c 18706 g 21683 t
ORIGIN

alignment_scores:
  Quality: 10.00      Length: 10
  Ratio: 1.000       Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AC017607/rev ..
Align seg 1/1 to reverse of: AC017607 from: 1 to: 81255
789 SerTyrAlaLysProLeuAsnLysGln 798
```

```

61164 TCTATGCAAGCCACTTAACAAAAGCAA 61135
|||||
seq_name: gb_ba:ECOTSF
seq_documentation_block:
LOCUS      ECOTSF                      91430 bp    DNA    linear    BCT 28-MAY-1999
DEFINITION Escherichia coli genomic DNA. (4.1 - 6.1 min).
VERSION    D83536
ACCESSION  D83536
KEYWORDS   Complete and shotgun sequencing; rpsB; tsf; smb; mukB; frr; rrf;
           yaec; cds; rnhB; yzzN; hlpA; skp; ompH; firA; fabZ; sefa;
           lpxA; lpxB; pgsB; dnaE; polC; accA; ldcC; lchH; yaeR; mesJ;
           yaeC; yaeB; yaeJ; cutF; nlpE; yaeF; proS; drpA; yaeB; rcsP;
           yaeC; abc; yaeD; yafB; yafC; yafD; yafE; dnrA; rnhA; rnh;
           dasF; herA; sdrA; dnaQ; mutD; yhhI; yafH; gnhA; lpcA; trfA; yafK;
           dinJ; yafI; thia; lafu; dlnP; yafN; yafO; yafP; prfH; pepD; gpt;
           gpp; yafA; cri; phoB; ompE; proB; proA; yfjZ; yfjY; yfjX; yfjT;
           yfjS; yfjQ; yfjP; dsdC; rocE; yagC; yagA; yjHh; yagB.
SOURCE     Escherichia coli (strain:K12) DNA.
ORGANISM   Escherichia coli
           Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
           Escherichia.
REFERENCE  1 (bases 1 to 91430)
AUTHORS   Musso,R., Di Lauro,R., Rosenberg,M. and de Crombrughe,B.
TITLE     Nucleotide sequence of the operator-promoter region of the
           galactose operon of Escherichia coli
JOURNAL   Proceedings of the National Academy of Sciences of the United
           States of America. 74 (1), 106-110 (1977)
MEDLINE   77102694
PUBMED    319453
REFERENCE  2 (bases 1 to 91430)
AUTHORS   Young,R.A. and Steitz,J.A.
TITLE     Complementary sequences 1700 nucleotides apart form a ribonuclease
           III cleavage site in Escherichia coli ribosomal precursor RNA
JOURNAL   Proceedings of the National Academy of Sciences of the United
           States of America. 75 (8), 3593-3597 (1978)
MEDLINE   79012454
PUBMED    358189
REFERENCE  3 (bases 1 to 91430)
AUTHORS   Mulligan,R.C. and Berg,P.
TITLE     Factors governing the expression of a bacterial gene in mammalian
           cells
JOURNAL   Molecular and cellular biology. 1 (5), 449-459 (1981)
MEDLINE   88094360
PUBMED    6100966
REFERENCE  4 (bases 1 to 91430)
AUTHORS   Little,J.W., Mount,D.W. and Yanisch-Perron,C.R.
TITLE     Purified lexA protein is a repressor of the recA and lexA genes
           of Escherichia coli
JOURNAL   Proceedings of the National Academy of Sciences of the United
           States of America. 78 (7), 4199-4203 (1981)
MEDLINE   82037806
PUBMED    7027255
REFERENCE  5 (bases 1 to 91430)
AUTHORS   An,G., Bendlak,D.S., Mameiak,L.A. and Friesen,J.D.
TITLE     Organization and nucleotide sequence of a new ribosomal operon in
           Escherichia coli containing the genes for ribosomal protein S2 and
           elongation factor Ts
JOURNAL   Nucleic acids research. 9 (16), 4163-4172 (1981)
MEDLINE   82059454
PUBMED    6272196
REFERENCE  6 (bases 1 to 91430)
AUTHORS   Overbeek,N., Bergmans,H., van Mansfeld,F. and Lugtenberg,B.
TITLE     Complete nucleotide sequence of phoE, the structural gene for the
           phosphate limitation inducible outer membrane pore protein of
           Escherichia coli K12
JOURNAL   Journal of molecular biology. 163 (4), 513-532 (1983)
MEDLINE   83189086
PUBMED    6341601
REFERENCE  7 (bases 1 to 91430)
AUTHORS   Kanaya,S. and Crouch,R.J.
TITLE     Low levels of RNase H activity in Escherichia coli PB2 rnh result
           from a single-base change in the structural gene of RNase H

```

---

```

JOURNAL    83185998
MEDLINE    6302075
PUBMED
REFERENCE  8 (bases 1 to 91430)
AUTHORS   Richardson,K.K., Foster,J. and Skopek,T.R.
TITLE     Nucleotide sequence of the xanthine guanine phosphoribosyl
           transferase gene of E. coli
JOURNAL    Nucleic acids research. 11 (24), 8809-8816 (1983)
MEDLINE    84169517
PUBMED
REFERENCE  9 (bases 1 to 91430)
AUTHORS   Pratt,D. and Subramani,S.
TITLE     Nucleotide sequence of the Escherichia coli xanthine-guanine
           phosphoribosyl transferase gene
JOURNAL    Nucleic acids research. 11 (24), 8817-8823 (1983)
MEDLINE    84169518
PUBMED    6324103
REFERENCE  10 (bases 1 to 91430)
AUTHORS   Deutch,A.H., Rushlow,K.E. and Smith,C.J.
TITLE     Analysis of the Escherichia coli proBA locus by DNA and protein
           sequencing
JOURNAL    Nucleic acids research. 12 (15), 6337-6355 (1984)
MEDLINE    84297232
PUBMED    6089111
REFERENCE  11 (bases 1 to 91430)
AUTHORS   Jagadeeswaran,P., Ashman,C.R., Roberts,S. and Langenberg,J.
TITLE     Nucleotide sequence and analysis of deletion mutants of the
           Escherichia coli gpt gene in plasmid pSV2 gpt
JOURNAL    Gene. 31 (1-3), 309-313 (1984)
MEDLINE    85128454
PUBMED    6396164
REFERENCE  12 (bases 1 to 91430)
AUTHORS   Nuesch,J. and Schumperli,D.
TITLE     Structural and functional organization of the gpt gene region of
           Escherichia coli
JOURNAL    Gene. 32 (1-2), 243-249 (1984)
MEDLINE    85155481
PUBMED    6397401
REFERENCE  13 (bases 1 to 91430)
AUTHORS   Becerril,B., Valle,F., Merino,E., Riba,L. and Bollivar,F.
TITLE     Repetitive extragenic palindromic (REP) sequences in the
           Escherichia coli gdhA gene
JOURNAL    Gene. 37 (1-3), 53-62 (1985)
MEDLINE    86031366
PUBMED    3902576
REFERENCE  14 (bases 1 to 91430)
AUTHORS   Nomura,T., Aiba,H. and Ishihama,A.
TITLE     Transcriptional organization of the convergent overlapping dnaQ-rnh
           genes of Escherichia coli
JOURNAL    The Journal of biological chemistry. 260 (11), 7122-7125 (1985)
MEDLINE    85207736
PUBMED    2987244
REFERENCE  15 (bases 1 to 91430)
AUTHORS   Icho,T., Sparrow,C.P. and Raetz,C.R.
TITLE     Molecular cloning and sequencing of the gene for CDP-diglyceride
           synthetase of Escherichia coli
JOURNAL    The Journal of biological chemistry. 260 (22), 12078-12083 (1985)
MEDLINE    86008268
PUBMED    2995358
REFERENCE  16 (bases 1 to 91430)
AUTHORS   Cox,E.C. and Horner,D.L.
TITLE     DNA sequence and coding properties of mutD(dnaQ) a dominant
           Escherichia coli mutator gene
JOURNAL    Journal of molecular biology. 190 (1), 113-117 (1986)
MEDLINE    87060973
PUBMED    3023634
REFERENCE  17 (bases 1 to 91430)
AUTHORS   Takano,K., Nakabeppu,Y., Maki,H., Horiuchi,T. and Sekiguchi,M.
TITLE     Structure and function of dnaQ and mutD mutators of Escherichia
           coli
JOURNAL    Molecular & general genetics : MGG. 205 (1), 9-13 (1986)
MEDLINE    87089093
PUBMED    3540531

```



```
/gene="fhuD"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="ferrichrome-iron transport protein FhuD"
/protein_id="AAB08582.1"
/db_xref="GI:1552730"
/translation="MSGPLISRRRLTAMALSPLLMQNTAHAAIDPNRIVALEWL
PEVLLALGIVGYVADTINYRLWSEPPDSDVIDGLRTEPNLELLTEMPSPFVW
SAGYGPSPMLARIAPGRGNFSDGKQPLAMARKSLTEADLLNLQSAARKLAQYED
FTRSMKPREVKKGARPLLTLTDPRHMLVFGPNSLFOELDEYGIIPNAWQGETNFWG
STAVSDRLAAVKDQVDFLCFDHNSDKMDALMATPLWQAMPFVRAGRFORVPAVFEY
ATLSAMHFRVLDNALGKA"
2537..4519
/gene="fhuB"
2537..4519
/gene="fhuB"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="ferrichrome-iron transport protein FhuB"
/protein_id="AAB08583.1"
/db_xref="GI:1552731"
/translation="MSKRIALPALLLIVIVATLTWMSQALPRSQWAAQANSP
DIDVLEQMTFHSILPRALISLLYGAGLGVLFQOVLNRNLEPTEITLGVATGAQLG
IYVITLAIPEGAWASQFAQAQACVVGLVFGVANGKRLSPVLLIAGLVSLYCGAI
NQLLVIFHDDQOSMFLSTGTQTQDGGVERLPQLGGVMLTLLRLPTLMGLD
DGVANGLASLARLAALSAIVISALVNAVGTIGTGFAPLLAKMLGARLLPR
LMASLICALIWLSDQITLWLTVMVSTGVTALIGAPLLMLPLRLRISAPDM
KYNRVAAERQHVAFALAGVLLMAMVVALSFGDAGHTWASGALLEDLMPWRP
RIMAAFAVGLAVAGCIQRLTGMPASPEVLGSSNAFGVLMFLFPGNAGWL
LPAGSLCAAVTILLIIMIAAGRGFSFHRMLAGMALSTMLLMQLQASGDPRAQV
LFWISGSTNATDAQWRTGIVMVILLAITPLCRRLTLPLGSDTARAVGMALTPTR
IALLLLAACLTATMTIGPLSFVGLMAPHAIARMGFRRTMETHIVISALVGLLILVFA
DWCGRMVFPPQIPAGLLSTFGAPYFIYLLKQSR"
complement(4677..5957)
/gene="gsa"
complement(4677..5957)
/gene="gsa"
/codon_start=1
/transl_table=11
/product="glutamine-1-semialdehyde aminotransferase"
/protein_id="AAB08584.1"
/db_xref="GI:1552732"
/translation="MSKENLYSAARELIPGVNSPVRFTGCVGTPLFIEKADGAYL
YDVKGKATIDYVSGMPVGLGHNHPAIRNAVIEAAERGLSFGAPTEMYKMAQLVTEL
VPTDMVRVNSGTATSAIRLARGFTGRDKIIFEGCYHGADCLLVKAGSGALT
GQPNSPGPVADPAKYLTCTYNDLASVRAAFQYPOETACIIVEPVAGNMNCVPLPE
PLPGLRLCDEFGALLIIDVMTGPRVALAGADYGVVVDLTCLGKIIGGMPYCAP
GGRDVMALAPGTQVYQAGTSLGPNPIMAGFACILNEVAQPGVHETIDELTRLAEG
LLEAAEEAGIPLVNVHVGGMFGIFFTDAESVTCYQDVMACDVERKRFHMMMLDEGVY
6182..7603
/Note="similar to voltage-gated chloride channel protein"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/protein_id="AAB08585.1"
/db_xref="GI:1552733"
/translation="NKTDTPSLETPOAARLRRLIROLLRDKTPTLALIFMAAVVGT
LVGLAAVDFDQVWALQNRMCALVHTADNYPVLLITVAFCLSAVLAMFGYFLVRYP
EAGSGSIEFEALQEDQPRVRWVPLPVKFFGGTGLGGVGLGREGPTVQIGNGVR
MYLDFRLKQDEARHTLLATGAAGAAAFNAPLAGILFIIEEMPRQYRLISIKVA
FIGVIMTYMTRINHEVALIDVGLSDAPLNTLWLYLILGIFIGFIPNFKWILGM
QDLHRVHGNTITKWLMMGAGIGGLGLGFVAPATSGGFNLPIATAGVNSMGMLV
FIFVARVITTLICFSSGAPGGIFAPMLALGTVLGTFAGMVAVELFPQVHLEAGTFAIA
GMGALLAASIRAPLTGIIIVLEMTDNYOLLIPMITTGLGATLLAQFTGGKPLYSAILA
RTLAQKAEQLARSAASENT"
7685..8029
/Note="hypothetical protein"
/codon_start=1
/transl_table=11
/evidence=not_experimental
```

```
/protein_id="AAB08586.1"
/db_xref="GI:1552734"
/translation="MSDDVALPLEFTDAANKVKSLIADENPNLKLRYVITGGCGSG
FOYGFTEDDQVNEGDMTIEKQGVGLVVDPMSLQYLVGGSDYDTEGLEGSRFIYNPNA
KSTCGGSSFSFI"
complement(8076..8699)
/Note="similar to Vibrio cholerae RecA protein"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/protein_id="AAB08587.1"
/db_xref="GI:1552735"
/translation="MLVYWLIDIVGTAVFAISGVLLAGKLRMDPFGVLVGLWVAVGGG
TIRDMALDHGPVFWKDPDLDLVVAMVTSMLTIVLRQPRRLPKWMLPVLDAVGLAVFV
GIGVNAKAFNAEAGPLIACMGVITGVGGGIIRDVLAREIPMLIRTIYATACIIGIV
HATAYTFVSPLETTASMMGMVTLIRLIRAAIRHRLKLPITFALDENG"
complement(8737..9537)
/gene="yadT"
complement(8737..9537)
/gene="yadT"
/Note="hypothetical"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/protein_id="AAB08588.1"
/db_xref="GI:1552736"
/translation="MAKSLFRALVALSFLAPLWLNAAPRVITLSPANTELAFAAGITP
VGVSSYDYPQQAQTEQVSTWQGMNLERIVAKPDLVIARWGNAEROVDOLASLGI
KVMVDATIEQIANALROLAPSPQDPKAEQAQASLLDQYAKLKAQYADKPKRVFL
QGINPPTSGRESTONQVLEVCGENIFKDSRVPWPQVSRQVLAQSQAIVITGSP
DQPKIKQWGEQLKPIVPLTSDMFERASPRILAAQQLCNALSOVD"
complement(9530..10228)
/Note="similar to purine nucleoside phosphorylase (dead)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/protein_id="AAB08589.1"
/db_xref="GI:1552737"
/translation="MKIGIIGAMEEVTLLRDKIENROTISLGCGEITYTGQNGTEVA
LLKSGIGKVAALGATLLEHCKKPDVINTGSAGGLAPTLKVGDIYVSEARYHDADV
TAFGEYQOLPGCPAGFKADDKLIAAAEACIAELNNAVRLIVSGDAFINGSVGLAK
IRHNPQATAVEMEATAIAHVCHNFVFPVVVRAISDVADQQSHLSFDEFLAAQAQS
SLWVSLQKLAHG"
10312..11829
/gene="dgt"
10312..11829
/gene="dgt"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="deoxyguanosine triphosphate"
/protein_id="AAB08590.1"
/db_xref="GI:1552738"
/translation="NAQIDFRKKINWHREYRSPQGVKTEHETLRFESDRGLINSPA
IRRLOQKQVFFLERNAAVRTRLTHSMEOVQVQGRYIAKEILSRLEKLLEAYGLDEL
TGPFESIVEMSCMHDPFEGHGAINDRQRHLPDEAESQPLDRCRVAAL
RLRDEEPLNELRKIRQDLFEQNAQGIIRLVHTLMRMLTWAQVGLTKYTRPAW
RGETPTHYLMKPKGYLSEAYTARLKEKLALYSRPLTWIMEAADDISYCVAD
LEDAVEKRIFTVEQLVHLHEAGQHEKGSLSFLVVENAWEKSRNSLSRSTEDQFFM
```

## alignment\_scores:

Quality:	10.00	Length:	10
Ratio:	1.000	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

## alignment\_block:

US-09-701-711-2 x ECU70214

Align seg 1/1 to: ECU70214 from: 1 to: 123171

648 ProPheTyrGluAsnPheTyrAlaGlyGly 657

|||||

30941 CCGTCTACGAGAACTCTATGCGGTGGT 30970



seq\_name: gb\_in:AC009366

seq\_documentation\_block:

LOCUS AC009366 169839 bp DNA linear INV 05-FEB-2002  
 DEFINITION Drosophila melanogaster 3L BAC RP98-48E10 (Roswell Park Cancer  
 Institute Drosophila BAC Library) complete sequence.  
 ACCESSION AC009366 AC010559 AC010717  
 VERSION AC009366.7 GI:18496988  
 KEYWORDS HTG:  
 SOURCE fruit fly.

## ORGANISM

Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

## REFERENCE

AUTHORS 1. (bases 1 to 169839)  
 Muzny, D., Scherer, S., Adams, M.D., Holt, R.A., Evans, C.A.,  
 Gocayne, J.D., Tabor, P., Williamson, A., Homs, F.H.,  
 Dugan-Rocha, S.D., Sodergren, E.S., Hodgson, A.H., Chen, R.C.,  
 Ayele, M., Scott, G.S., Worley, K.W., Anamatiades, P.G., Brandon, R.C.,  
 Rogers, Y., An, H., Baldwin, D., Beeson, K.Y., Brown, M., Buhay, C.,  
 Busam, D.A., Center, A., Chen, G., Chen, Z., Clerc-Blankenburg, K.,  
 Davenport, L.B., Dietz, S.M., Ding, Y., Dodson, K., Doup, L.E.,  
 Draper, H., Emery-Cohen, A., Ferreira, S., Garg, N.D.S., Houck, J.,  
 Hostin, D., Howland, T.J., Hume, J., Ibegwam, C., Jalali, M., Kovar, C.,  
 Liu, W., Mattei, B., McIntosh, T.C., Morgan, M., Moy, M., Murphy, B.,  
 Nelson, K.A., Ndassa, Y., Nguyen, N., Perez, L., Pittman, G.S., Puri, V.,  
 Scheeler, F., Shen, H., Strong, R., Tector, C., Wang, Q., Williams, S.M.,  
 Wheeler, J., Zaveri, J.S., Zhou, J., Zorrilla, S., Smith, H.O.,  
 Xiang, D., Weinstock, G., Gibbs, R. and Venter, J.C.

## Direct Submission

## Unpublished

## 2 (bases 1 to 169839)

Worley, K.C., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,  
 Alsdrooks, S.L., Amarantunge, H.C., Are, J.R., Banks, T., Barbara, J.,  
 Benton, J., Briane, K., Blankenburg, K., Bonnin, D., Bouck, J.,  
 Bowle, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,  
 Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F.,  
 Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,  
 Chen, Z., Choudhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,  
 Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,  
 Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,  
 Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,  
 Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,  
 Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,  
 Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,  
 Garza, N., Gill, R., Correll, J.H., Guevara, W., Gunaratne, P., Hale, S.,  
 Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,  
 Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C.,  
 Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,  
 Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,  
 Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,  
 Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,  
 Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,  
 Loulsegue, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,  
 Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,  
 Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M.,  
 Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,  
 Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,  
 Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S.,  
 Ogum, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,  
 Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,  
 Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M.,  
 Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoostari, N.,  
 Sisson, I., Sodergren, E., Sonake, T., Sparks, A., Stanley, H.,  
 Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,  
 Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,  
 Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R.,  
 Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,  
 Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S.,  
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstock, G. and Gibbs, R.

## Direct Submission

## TITLE

## JOURNAL

REFERENCE 3 (bases 1 to 169839)

AUTHORS Worley, K.C.

TITLE Direct Submission

## JOURNAL

Submitted (20-AUG-1999) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

## REFERENCE

AUTHORS 4 (bases 1 to 169839)

TITLE BCM-HGSC.

## JOURNAL

Submitted (05-FEB-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

## COMMENT

On Feb 5, 2002 this sequence version replaced gi:6056147 gi:6056117  
 gi:5916439  
 INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email  
 gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the  
 entire insert of this clone. Overlapping regions of clones are only  
 sequenced and submitted once, so the sequence for the remainder of  
 the insert may be found in the record for the adjacent clones.  
 Overlapping clones are noted at the beginning and end of the  
 Features listing.

## ANNOTATION OF FEATURES:

- STSS are identified using ePCR (Genome Res. 7:541-550) searches  
 of a local database that includes entries from dbSTS, GDB, and  
 local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,  
 unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST  
 (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the  
 EST and cDNA sequences. Genes demonstrate at least two exons  
 flanked by consensus splice sites that maintained sequence  
 continuity across the splice junctions. Sequences that are not  
 identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum  
 standard of double strand coverage with a minimum of 2 clones and 2  
 reads with no ambiguities or 2 chemistries with a minimum of 2  
 clones and 3 reads with no ambiguities. If the sequence quality for  
 a region does not meet this standard, it will be indicated in the  
 annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality  
 standards - estimated error rate less than 1 per 10,000 bases.  
 Reports of lowest quality individual bases and measures of base  
 quality are listed below. Description of the metrics can be found  
 at URL: http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.  
 html.

## FEATURES

Location/Qualifiers

1..169839

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/chromosome="3L"

/clone="RP98-48E10"

BASE COUNT 44884 a 39408 c 39261 g 46286 t

## ORIGIN

## alignment\_scores:

Quality: 10.00 Length: 10

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-701-711-2 x AC009366

Align seg 1/1 to: AC009366 from: 1 to: 169839

789 SerTyrAlaLysProLeuAnLysLysGln 798

|||||

99653 TCATTATGCAAGCCACTTACAAAAAGCAA 99682

seq\_name: gb\_in:AC023676

seq\_documentation\_block:

LOCUS AC023676 179611 bp DNA linear INV 03-JAN-2002  
DEFINITION Drosophila melanogaster 3L BAC RP98-22C11 (Roswell Park Cancer  
Institute Drosophila BAC Library) complete sequence.

ACCESSION AC023676

VERSION AC023676.3 GI:17860983

KEYWORDS HTC

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

## REFERENCE

## AUTHORS

Muzny,D., Scherer,S., Adams,M.D., Holt,R.A., Evans,C.A.,  
Gocayne,J.D., Tabor,P., Williamson,A., Homsí,F.H.,  
Dugan-Rocha,S.D., Sodergren,E.S., Hodgson,A.H., Chen,R.C.,  
Ayele,M.A., Scott,G.S., Worley,K.W., Amaratunga,H.C.,  
Benton,J., Bimaga,K., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,  
Busam,D.A., Center,A., Chen,G., Chen,Z., Clerc-Blankenburg,K.,  
Davenport,L.B., Dietz,S.M., Ding,Y., Dodson,K., Doup,L.E.,  
Draper,H., Emery-Cohen,A., Ferreira,S., Garg,N.D.S., Houck,J.,  
Hostin,D., Howland,T.J., Hume,J., Ibegwam,C., Jalali,M., Kovar,C.,  
Liu,W., Mattei,B., McIntosh,T.C., Morgan,M., Moy,M., Murphy,B.,  
Nelson,K.A., Ndassa,Y., Nguyen,N., Perez,L., Pittman,G.S., Puri,V.,  
Schaefer,F., Shen,H., Strong,R., Tector,C., Wang,Q., Williams,S.M.,  
Xiang,J., Zaveri,J.S., Zhou,J., Zorrilla,S., Smith,H.O.,  
Wheeler,D., Weinstock,G., Gibbs,R. and Venter,J.C.

Direct Submission

Unpublished

2 (bases 1 to 179611)

Worley,K.C.

Direct Submission

Submitted (17-FEB-2000) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 179611)

Worley,K.C., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbaria,J.,  
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,  
Bowie,S., Brieval,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,  
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,  
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,  
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,  
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,  
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,  
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,  
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,  
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,  
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,  
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,  
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,  
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,  
Hollins,B., Homsí,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,  
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,  
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,  
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Lucier,A., Luna,R.,  
Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,  
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,  
Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,  
Mel,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,  
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,  
Nguyen,A., Nguyen,G., Oragunye,N., Nickerson,E., Nwokenkwo,S.,  
Ogulu,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,  
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,  
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,  
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N.,  
Slison,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H.,  
Ston,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,  
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,  
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,  
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,  
Watlington,S., Williams,G., Williamson,A., Wleczky,R., Wooden,S.,  
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
Weinstock,G. and Gibbs,R.

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

Submitted (29-DEC-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

5 (bases 1 to 179611)

Worley,K.C., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbaria,J.,  
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,  
Bowie,S., Brieval,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,  
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,  
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,  
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,  
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,  
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,  
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,  
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,  
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,  
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,  
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,  
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,  
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,  
Hollins,B., Homsí,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,  
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,  
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,  
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Lucier,A., Luna,R.,  
Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,  
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,  
Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,  
Mel,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,  
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,  
Nguyen,A., Nguyen,G., Oragunye,N., Nickerson,E., Nwokenkwo,S.,  
Ogulu,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,  
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,  
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,  
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N.,  
Slison,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H.,  
Ston,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,  
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,  
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,  
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,  
Watlington,S., Williams,G., Williamson,A., Wleczky,R., Wooden,S.,  
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
Weinstock,G. and Gibbs,R.

Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovacs, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, B.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseghe, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshitari, N., Sisson, I., Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G., and Gibbs, R.

Direct Submission  
Submitted (03-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Dec 15, 2001 this sequence version replaced gi:6996769.  
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email [gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)

TITLE  
JOURNAL  
COMMENT

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the features listing.

#### ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.  
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.  
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base annotation as Low Coverage.

alignment\_scores:  
Quality: 10.00 Length: 10  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

#### alignment\_block:

US-09-701-711-2 x AC023676 ..

Align seg 1/1 to: AC023676 from: 1 to: 179611

789 SerTyrAlaLysProLeuAsnLysLysGln 798

|||||

53183 TCTTATGCAAGCCACTTAACAAAGCA 53212

seq\_name: gb\_ba:AJ414154  
seq\_documentation\_block:  
LOCUS AJ414154 204050 bp DNA linear BCT 04-OCT-2001  
DEFINITION Yersinia pestis strain CO92 complete genome; segment 14/20.  
ACCESSION AJ414154 AL590842  
VERSION AJ414154.1 GI:15980810  
KEYWORDS Yersinia pestis.  
SOURCE Yersinia pestis.  
ORGANISM Yersinia pestis  
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Yersinia.  
REFERENCE 1 (bases 1 to 204050)  
AUTHORS Parkhill, J., Wren, B.W., Thomson, N.R., Titball, R.W., Holden, M.T.G., Prentice, M.B., Sebaihia, M., James, K.D., Churcher, C., Mungall, K.L., Baker, S., Basham, D., Bentley, S.D., Brooks, K., Cerdono-Tarraga, A.M., Chillingworth, T., Cronin, A., Davies, R.M., Davis, P., Dougan, G., Feltwell, T., Hamlin, N., Holroyd, S., Jagels, K., Leather, S., Karlyshev, A.V., Moule, S., Oyston, P.C.F., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., Whitehead, S. and Barrell, B.G.  
TITLE Genome sequence of Yersinia pestis, the causative agent of plague  
JOURNAL Nature 413 (6855), 523-527 (2001)  
MEDLINE 21470413  
REFERENCE 2 (bases 1 to 204050)  
AUTHORS Parkhill, J.  
TITLE Direct Submission  
JOURNAL Submitted (04-OCT-2001) Submitted on behalf of the Yersinia sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: [parkhill@sanger.ac.uk](mailto:parkhill@sanger.ac.uk)  
COMMENT Notes:  
Details of Y. pestis sequencing at the Sanger Centre are available on the World Wide Web.  
(URL, [http://www.sanger.ac.uk/Projects/Y\\_pestis/](http://www.sanger.ac.uk/Projects/Y_pestis/)).  
FEATURES  
source  
1. 204050  
/organism="Yersinia pestis"  
/strain="CO92"  
/db\_xref="taxon:632"  
/note="biovar: Orientalis"  
complement(247..1095)  
/gene="YPO2841"  
complement(247..1095)  
/gene="YPO2841"  
/note="Similar to Agrobacterium tumefaciens mannopine permealase MotD TR:Q44381 (EMBL:AF242881) (273 aa) fasta scores: E(): 0, 41.5% id in 258 aa, and to Pseudomonas aeruginosa probable ABC transporter permealase PA0606 TR:AAG03995 (EMBL:AF004497) (276 aa) fasta scores: E(): 0, 65.4% id in 260 aa"  
/codon\_start=1  
/transl\_table=1  
/product="putative ABC transporter permealase"  
/protein\_id="CAC92093.1"  
/db\_xref="GI:15980811"  
/translation="MPKPQKALKQKQAWLLRGKWFNFYSAAILLFLIVPLVIVPL SFNAGSPLSYPMAGFSURVQTPNSDWLGNLSLLIAPLATLLATGLGLVLAIGL VRGEFGKALVMAWLIISPMVAPVVIAGVGFEEFARLSNLSYGLVLAHALICGVFV VITVTAVLKSYDQNLSSRAASLGSASQWRVFKVTLPLIAPGVFSGALFAPAFVSDEVV VTFLFASPRQRTLPKRMFAGIRENDPTIAAASLMVGASLLILLIIVELLRRSRRLR HHPVSP"  
complement(292..357)  
/gene="YPO2841"  
/note="One of 6 probable transmembrane helices predicted for YPO2841 by TMHMM2.0"  
complement(376..597)  
/gene="YPO2841"  
/note="Ypfam match to entry PF00528 BPD\_transp, Binding-protein-dependent transport systems inner membrane component, score 59.50, E-value 7.3e-14"  
complement(454..519)  
/gene="YPO2841"  
/note="One of 6 probable transmembrane helices predicted

```
misc_feature
for YPO2841 by TMHMM2.0"
complement(508. .594)
/gene="YPO2841"
/Note="PS00402 Binding-protein-dependent transport systems
inner membrane comp. sign."
complement(583. .648)
/gene="YPO2841"
/Note="one of 6 probable transmembrane helices predicted
for YPO2841 by TMHMM2.0"
complement(694. .759)
/gene="YPO2841"
/Note="one of 6 probable transmembrane helices predicted
for YPO2841 by TMHMM2.0"
complement(787. .852)
/gene="YPO2841"
/Note="one of 6 probable transmembrane helices predicted
for YPO2841 by TMHMM2.0"
complement(967. .1032)
/gene="YPO2841"
/Note="one of 6 probable transmembrane helices predicted
for YPO2841 by TMHMM2.0"
complement(1157. .2422)
/gene="YPO2842"
/Note="one of 6 probable transmembrane helices predicted
for YPO2842"
complement(1157. .2422)
/Note="Similarity to the N-terminus of Agrobacterium
tumefaciens mannopine permease MotC TR:Q44382
(EMBL:AF242881) (294 aa) fasta scores: E(): 2.7e-29, 41.3%
id in 213 aa, and to the full length Pseudomonas
aeruginosa probable ABC transporter permease PA0605
TR:AA03994 (EMBL:AE004497) (415 aa) fasta scores: E(): 0,
54.7% id in 397 aa"
/codon_start=1
/transl_table=11
/product="putative ABC transporter permease"
/protein_id="CAC92094.1"
/db_xref="GI:15980812"
/translation="MSQEMVTKPVADERQSGPILRQLRKVKAAAYRRSLLITPL
LLFIVSFIFATILGKSIDNPELRVAMPVTVAVLANPGDALPDEAMFALITDLR
NARSLGMLSTAKMSYEDNRYSRLIMSTLRKVPAGPVREALISALPLWGLSTWQS
LQRTQPLTRYLLLSAFDRKVDVSGOIKALPAQOALYVNLRLTLNMAAVYLLCVT
LGYPLAYWLAKQPTGRANLLMLVLLPWTSLIVRTASWIVLLSGSLNLSLGSRL
TEQPLALVFNAGVYISITHILLPPLVILFVAVMKGISPNVRAISLGAHPFFAFWR
VYVPTQNGVTAALLFEMPAIGYYITPALLGGPDQMVSYFVAFPTNTTNWGMAAA
LGSLLIIVLLVYVYIRVTRTQAEATH"
complement(1196. .1261)
/gene="YPO2842"
/Note="one of 6 probable transmembrane helices predicted
for YPO2842 by TMHMM2.0"
complement(1277. .1498)
/gene="YPO2842"
/Note="Pfam match to entry PF00528 BPD_transp,
Binding-protein-dependent transport systems inner membrane
component, score 40.70, E-value 3.2e-08"
complement(1328. .1393)
/gene="YPO2842"
/Note="one of 6 probable transmembrane helices predicted
for YPO2842 by TMHMM2.0"
complement(1409. .1495)
/gene="YPO2842"
/Note="PS00402 Binding-protein-dependent transport systems
inner membrane comp. sign."
complement(1493. .1558)
/gene="YPO2842"
/Note="one of 6 probable transmembrane helices predicted
for YPO2842 by TMHMM2.0"
complement(1652. .1717)
/gene="YPO2842"
/Note="one of 6 probable transmembrane helices predicted
for YPO2842 by TMHMM2.0"
complement(1739. .1804)
/gene="YPO2842"
/Note="one of 6 probable transmembrane helices predicted
```

## misc\_feature

```
for YPO2842 by TMHMM2.0"
complement(2246. .2311)
/gene="YPO2842"
/Note="one of 6 probable transmembrane helices predicted
for YPO2842 by TMHMM2.0"
complement(2469. .3497)
/gene="YPO2843"
complement(2469. .3497)
/gene="YPO2843"
/Note="Similar to Agrobacterium radiobacter agropine
permease protein AgtB TR:Q9R701 (EMBL:AF242881) (364 aa)
fasta scores: E(): 9.9e-26, 30.3% id in 320 aa, and to
Pseudomonas aeruginosa hypothetical protein PA0222
TR:AA03611 (EMBL:AE004460) (352 aa) fasta scores: E(): 0,
56.1% id in 342 aa"
/codon_start=1
/transl_table=11
/product="putative transport system periplasmic protein"
/protein_id="CAC92095.1"
/db_xref="GI:15980813"
/translation="MFKIMTLTLLAGWVFQTHAETLTIVVSFGGLNKDAQVKAFYKPF
MAACKGTIEAGYNGEMARIRAMVETGQIGWDIVEGPELLRGNEGLFIDWSKL
GDERQFINGTVSQCGSLFGLSMILLIYDASRKTPTVTSWADFNNVTTFGRSLRKRTA
KFTMEIALLADGVKREDYVRLATPEGVRFNKLDQKPNQWWSGAQPLQWLVSFG
DVMGVSYNGRVGSALKEGHDVRFMYMTDSIYDMSWTIVKSKKVKLPAQFTAFANLA
ENQKVAENIAYGPTNINATLMIDPEIAANLPTAPANLAEAFPMNQFVWEHSEDLQ
RFSNWAAR"
complement(3550. .4839)
/gene="goag"
/Note="goag"
/Note="gabt; YPO2844"
complement(3550. .4839)
/gene="YPO2844"
/EC_number="2.6.1.19"
/Note="Similar to Escherichia coli 4-aminobutyrate
aminotransferase GoG SW:GOAG_ECOLI (P50457) (421 aa)
fasta scores: E(): 0, 65.7% id in 428 aa, and to
Pseudomonas aeruginosa 4-aminobutyrate aminotransferase
Gabt TR:AA03655 (EMBL:AE004465) (426 aa) fasta scores:
E(): 0, 53.0% id in 423 aa"
/codon_start=1
/transl_table=11
/product="putative aminobutyrate aminotransferase"
/protein_id="CAC92096.1"
/db_xref="GI:15980814"
/translation="MSHRELQRKEATPRGIGVLCDYAVVAENATLWDEQGREYID
FTAGLATINIGHRHPKVMAAVQRQDDQFTHAYQVVPVAVVTLAEKINSLAPISDSN
MTAGNSKTAFTFTGVEAIEANVKIARAATGPRGPIAFSGAFHGTLLAMALTGRAVP
YKVGPPFPASIFHALYENELYGVSVEAISVERLFRCDISPTQVAILPEPIQEGG
GFNIAPPFVSALRTLCDEHGILLIADDEVQTFARTGKLFAMEYYPDTKVDVITWAKS
LGGMPISAVTGRADMDAPLPGSLGGTYAGNPLAAVASLAVLDIIAEKLCERALLI
GAKLVLEKQAQMSNAIIVGRAGSMVAVFNDFPVSGKPSPELTPAYRQALEBGLL
LLSCGVHSNVRFLYPLTIPDKQFKOAMNILLRLAS"
complement(3613. .4749)
/gene="goag"
/Note="Pfam match to entry PF00202 aminotran_3,
Aminotransferases class-III pyridoxal-phosphate, score
459.70, E-value 2.4e-159"
complement(4000. .4116)
```

## misc\_feature

alignment\_scores:  
Quality: 10.00 Length: 10  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-701-711-2 x AJ414154

Align seg 1/1 to: AJ414154 from: 1 to: 204050

599 GlnGlyMetSerHisSerValAspLeuThr 608  
|||||  
90253 CAAGGAATGAGTCACCTCTGTGACCTGACT 90282

```
seq_name: gb_ba:AJ414146
seq_documentation_block:
LOCUS AJ414146 210050 bp DNA linear BCT 04-OCT-2001
DEFINITION Yersinia pestis strain C092 complete genome; segment 6/20.
ACCESSION AJ414146 AL590842
VERSION AJ414146.1 GI:15979072
KEYWORDS
SOURCE
ORGANISM Yersinia pestis.
Yersinia pestis.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Yersinia.
1 (bases 1 to 210050)
Parkhill,J., Wren,B.W., Thomson,N.R., Titball,R.W., Holden,M.T.G.,
Prentice,M.B., Sebahia,M., James,K.D., Churcher,C., Mungall,K.L.,
Baker,S., Basham,D., Bentley,S.D., Brooks,K., Cerdeno-Tarraga,A.M.,
Chillingworth,T., Cronin,A., Davies,R.M., Davis,P., Dougan,G.,
Feltwell,I., Hamlin,N., Holroyd,S., Jagels,K., Leather,S.,
Karlshede,A.V., Moule,S., Oyston,P.C.F., Quail,M., Rutherford,K.,
Simmonds,M., Skelton,J., Stevens,K., Whitehead,S. and Barrrell,B.G.
Genome sequence of Yersinia pestis, the causative agent of plague
Nature 413 (6855), 523-527 (2001)
21470413
2 (bases 1 to 210050)
Parkhill,J.
Direct Submission
Submitted (04-OCT-2001) Submitted on behalf of the Yersinia
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
Notes:
Details of Y. pestis sequencing at the Sanger Centre are available
on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/Y_pestis/).
FEATURES             source
1. 210050
/organism="Yersinia pestis"
/strain="C092"
/db_xref="taxon:632"
/notes="biovar: Orientalis"
829..11946
/genes="yapH"
/genes="YpO1004"
829..11946
/genes="yapH"
/notes="Previously sequenced as Yersinia pestis YapH
protein TR:CAC14227 (EMBL:AJ277631) (3705 aa)
scores: E(): 0, 99.6% id in 3705 aa. Similar in regions to
Escherichia coli Ycha protein TR:09JMS3 (EMBL:AP001918)
(1371 aa) fasta scores: E(): 0, 39.1% id in 1405 aa, and
to Escherichia coli adhesin Aida-I precursor SW:AIDA.ECOLI
(Q03155) (1286 aa) fasta scores: E(): 0, 29.2% id in 1286
aa. Contains a possible N-terminal signal sequence."
/codon_start=1
/transl_table=11
/product="putative autotransporter protein"
/protein_id="CAC89847.1"
/db_xref="GI:15979073"
/translations="MTIPKVIWNASLNVVVVSELAQRIKTKSRNLISEGVLPKF
EQSWKLFKRNLLSLGSLVFLSGPFAADITVSTQELSAALNSNGIYKIIIGA
DITLIGSLFNMVTSQVYIDGGQFGLTVNNTTNYGLVSSGSGTLTLQNMKSIDNSA
YYSMVYLNAGTAVNYIYNNIDFLGSQILYMGAYATNSIMTFGDIINDVYVNDRA
QETGVNKLAFTAFGRVHTHTGSSVTSVSTGGANNTWDFASGADVKTIDRTSGDL
TSPGVNAFTPADGASFELIANQVPSGTTNRRGLEIGSYNSIDFGSGGVILQSR
SDGSIISGNGIDATNNAAGINNAGSDANVYINLGTGSLKATNTGILATKNNAS
DIYRAGDTATATGTSATHNGTGVKIRKNDGTFITTAGIASSEIKSEISVDNTDG
TIPATAGTGVNVLASAILNFGGTINTSFANGTITFAGTGHTLDTLNLGTLGIA
LSNVAGNLTLSVNTLNGTALNSLTGLTVDSLNGRNTINIEGAGIGIAATNEL
NTPDACALDINVGAGTIGQATGGVNLASNLINVAATLGTALQITDGDINTTITIG
NETQLAENATAIYELGSSSKTLNNNGTIKGSVIFAGVADHIIINNGTLDGLTTGAG
SLDAGTGLNTLNSDELAAATSLQGFNTINLVDSTHITVSDNDIGSGMWNIDSS
ELDFGTFGLIHATLGAGTSAIVNNSANVLEQAFMGATQVQVNOGALTSANQ
LGSAKIGLDLTLLNFHVLNGTLNVAKNLATTAFDFGTVGGAFSGIVNLT
```

```
DLLSARETVKTAENQDFSKWLSQSPKWSVQVRIEIPERYAAAVAKOYHALENIYDPKL
AAELAANGTGDVDANRIVGKRINDELGMELDMALTHEVLSAKGASSLLDNLWMEYLI
SP"
/misc_feature
  complement(13099. .13164)
  /gene="yPO1005"
  /note="Pfam match to entry PF00560 LRR, Leucine Rich
Repeat, score 9.30, E-value 44"
  complement(13219. .13290)
  /gene="yPO1005"
  /note="Pfam match to entry PF00560 LRR, Leucine Rich
Repeat, score 12.80, E-value 8.4"
  complement(13294. .13353)
  /gene="yPO1005"
  /note="Pfam match to entry PF00560 LRR, Leucine Rich
Repeat, score 10.40, E-value 30"
  complement(13468. .13539)
  /gene="yPO1005"
  /note="Pfam match to entry PF00560 LRR, Leucine Rich
Repeat, score 17.30, E-value 0.36"
  complement(13543. .13602)
  /gene="yPO1005"
  /note="Pfam match to entry PF00560 LRR, Leucine Rich
Repeat, score 12.90, E-value 7.8"
  complement(14492. .15367)
  /gene="yPO1006"
  complement(14492. .15367)
  /gene="yPO1006"
  /note="Similar to neighbouring CDSS YPO1005 and YPO1007.
Similar to regions of Salmonella typhimurium secreted
protein h2 TR:O9RPH0 (EMBL:AF160727) (788 aa) fasta
scores: E(): 2.5e-24, 41.9% id in 272 aa, and to regions
of Yersinia pestis outer membrane protein YopM TR:O68701
(EMBL:AF053946) (409 aa) fasta scores: E(): 7e-23, 40.5%
id in 259 aa"
  /codon_start=1
  /transl_table=11
  /product="putative antigenic leucine-rich repeat protein"
  /protein_id="CAC89849.1"
  /db_xref="GI:15979075"
  /translation="MYLSNITSVSMNPNGIPDREIHADRPAAATALTADYHAIWEKWE
NDPRTVAGEQGAOAVARMKECLLENNAERLNLSLDTLSLPTLPNCELNIICNLTE
LPTLPDNLQTLKASYNOLRTLPTLTPASLSLSKVHMELEKPELPGLKTLQVGC
NTSLQPLPVLPSLESDISNCLTEPLTPNSLKELDAHGQLRLDPLPISLRL
NWAYNQLALPENLPGSLRCIYTEYNQLSQLPDLAHLRQNCNCLDGNPLSPSTLLAL
LRLSTKPNYQGPRI"
  complement(14657. .14728)
  /gene="yPO1006"
  /note="Pfam match to entry PF00560 LRR, Leucine Rich
Repeat, score 15.40, E-value 1.3"
  complement(14732. .14791)
  /gene="yPO1006"
  /note="Pfam match to entry PF00560 LRR, Leucine Rich
Repeat, score 9.30, E-value 43"
  complement(14981. .15040)
  /gene="yPO1006"
  /note="Pfam match to entry PF00560 LRR, Leucine Rich
Repeat, score 11.30, E-value 22"
  alignment_scores:
    Quality: 10.00      Length: 10
    Ratio: 1.000        Gaps: 0
    Percent Similarity: 100.000      Percent Identity: 100.000
  alignment_block:
    US-09-701-711-2 x AJ414146
    Align seq 1/1 to: AJ414146 from: 1 to: 210050
    648 ProPheTyrGluAsnPheTyrAlaGlyGly 657
    |||||
    76307 CCATCTACGAAACTCTATGACGGTGT 76336
  seq_name: gb_ba:AL627266
```

```
seq_documentation_block:
LOCUS      AL627266                268050 bp      DNA      linear      BCT 25-OCT-2001
DEFINITION      Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
complete chromosome, segment 2/20.
ACCESSION      AL627266
VERSION        AL513382
KEYWORDS        AL627266.1 GI:16501496
SOURCE          Salmonella enterica subsp. enterica serovar Typhi.
ORGANISM        Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Salmonella.
REFERENCE      1 (bases 1 to 268050)
AUTHORS        Parkhill, J., Dougan, G., James, K.D., Thomson, N.R., Pickard, D.,
Wain, J., Church, C., Mungall, K.L., Bentley, S.D., Holden, M.T.G.,
Sebaihia, M., Baker, S., Basham, D., Brooks, K., Chillingworth, T.,
Conerton, P., Cronin, A., Davis, P., Davies, R.M., Dowd, L., White, N.,
Farrar, J., Felwell, T., Hamlin, N., Haque, A., Hien, T.T., Holroyd, S.,
Jagels, K., Krogh, A., Larsen, T.S., Leather, S., Moule, S., O'Gaora, P.,
Parry, C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J.,
Stevens, K., Whitehead, S. and Barrell, B.G.
Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18
NATURE 413 (6858), 848-852 (2001)
JOURNAL        11677608
PUBMED
REFERENCE      2 (bases 1 to 268050)
AUTHORS        Parkhill, J.
TITLE          Direct Submission
JOURNAL        Submitted (25-OCT-2001) Submitted on behalf of the Salmonella
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, UK
COMMENT        E-mail: parkhill@sanger.ac.uk
Notes:
Details of S. typhi sequencing at the Sanger Centre are available
on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/S_typhi/).
FEATURES
source
  1. .268050
  Location/Qualifiers
    /organism="Salmonella enterica subsp. enterica serovar
Typhi"
    /strain="CT18"
    /db_xref="taxon:90370"
    181. .906
    /gene="STY0239"
    /note="rpsB"
    181. .906
    /gene="STY0239"
    /note="Orthologue of E. coli rpsB (RS2_ECOLI); Pasta hit
to RS2_ECOLI (240 aa), 98% identity in 240 aa overlap"
    /codon_start=1
    /transl_table=11
    /product="30S ribosomal protein S2"
    /protein_id="CAD08674.1"
    /db_xref="GI:16501497"
    /translation="MATVSMRDLKAGVHFHQTRYWNPMPFFGARKKVIINLE
KTVPMFEALAEINLKISARKKILFVGTAKRAAEAVKAANSCDQFFVNRHWGGMLT
NWKTVRSIKRLKDLQETQSDGTQFEKLTKEALMRTELEKLSIGIKDGLPDA
LFVIDADHEHIAIKENANLIGIPVFAIVDTNSDPDGVDFVPGNDDAIRAIVSLYLGAVA
ATVREGRSQDLASQAESEFVEAE"
    196. .231
    /gene="STY0239"
    /note="PS00962 Ribosomal protein S2 signature 1"
    205. .855
    /gene="STY0239"
    /note="Pfam match to entry PF00318 Ribosomal_S2, Ribosomal
protein S2, score 441.70, E-value 6.5e-129"
    652. .726
    /gene="STY0239"
    /note="PS00963 Ribosomal protein S2 signature 2"
    1164. .2015
    /gene="STY0240"
    /note="tsf"
    1164. .2015
  misc_feature
  misc_feature
  misc_feature
  gene
  CDS
```

```

/misc_feature
/EC_number=1.1.1.1.-"
/notes="Orthologue of E. coli tsf (EFTS_ECOLI); Fasta hit to EFTS_ECOLI (282 aa), 97% identity in 282 aa overlap"
/codon_start=1
/transl_table=11
/product="elongation factor Ts"
/protein_id="CAD08675.1"
/db_xref="GI:16501498"
/translation="WAEITASLVKELRERTGAGMDCKKALTEANGDIELATENMRKS
GAIKAAKAGNVAADGVIKTKIDGNVAFILEVNCQTDFAKAGQAFADKVLDRAYA
GKITDEVILKAGFEERVALVAKIGENINIRRVASLEGVLGYSYQHGARGVILRAAK
ADEELVKQAMHVAASKPEFVEDVADVKEYOQOLDIAMQSKPKKEIAEKVMVEG
RNMKKTGVSLSGQCFVMEPSKSVGOLLKEHNADVGTGIRFEVGEVIEKVERDFAEV
AAMSQS"
1170..1289
/notes="STY0240"
/notes="Pfam match to entry PF02094 TS-N, TS-N domain, score 81.00, E-value 2.4e-20"
1197..1244
/notes="STY0240"
/notes="PS01126 Elongation factor Ts signature 1"
1332..1952
/notes="STY0240"
/notes="Pfam match to entry PF00889 EF-TS, Elongation factor TS, score 356.50, E-value 2.9e-103"
1386..1418
/notes="STY0240"
/notes="PS01127 Elongation factor Ts signature 2"
2160..2885
/notes="STY0241"
/notes="pyrH"
2160..2885
/notes="STY0241"
/notes="Orthologue of E. coli pyrH (PYRH_ECOLI); Fasta hit to PYRH_ECOLI (240 aa), 98% identity in 240 aa overlap"
/codon_start=1
/transl_table=11
/product="uridine 5'-monophosphate kinase"
/protein_id="CAD08676.1"
/db_xref="GI:16501499"
/translation="WATNAKPVYKRIKLLKSGEALQGTGEGIDASILRMAQEIHEL
VFLGTQGVVIGGNLFRGAGLAKAGMNVGDHGMGLATVMNGLAMRDALHRAVYNA
RLMSAIPNGVCDNTSWAEASILLRNRRVYLSAGTGNPFPTDSACILRGIEIADV
VLKATKDVGVFTADPAKDPASATMYDQLTYSVLDKELKVMDLAAEFLARDHKLPVRF
NNKPCALRRVVMGEKESGLITE"
2187..2738
/notes="STY0241"
/notes="Pfam match to entry PF00696 aakinasase, Amino acid kinase family, score 153.90, E-value 2.9e-42"
3032..3389
/notes="STY0242"
/notes="frr"
3032..3389
/notes="STY0242"
/notes="Orthologue of E. coli frr (RFR_ECOLI); Fasta hit to RFR_ECOLI (185 aa), 94% identity in 185 aa overlap"
/codon_start=1
/transl_table=11
/product="ribosome recycling factor"
/protein_id="CAD08677.1"
/db_xref="GI:16501500"
/translation="MISDIRKDAEVRMEKVEAFKTOISKVRTGRASPSSLIDGIVVEY
YGTPTPLROLASTVEDSPTLKNVDFRSMGPAVEKAIMASDLGLNPPSSAGDIDRVPL
PLTEERRDLKPIVRGEAEQARVAVNRVRRANDKVKALLDKDKAISSEDDRRSQEEV
QKMTDAIKKVDALADKAEELMQF"
3086..3380
/notes="STY0242"
/notes="Pfam match to entry PF01765 RRF, Ribosome recycling factor, score 339.90, E-value 2.9e-98"
3731..4927
/notes="dxr"
/notes="STY0243; yaem"
3731..4927

```

---

```

/misc_feature
/EC_number=1.1.1.1.-"
/notes="Similar to Escherichia coli 1-deoxy-D-xylulose 5-phosphate reductoisomerase dxr SW:DXR_ECOLI (P45568; P77209) (398 aa) fasta scores: E): 0, 88.9% id in 397 aa Orthologue of E. coli yaem (DXR_ECOLI); Fasta hit to DXR_ECOLI (398 aa), 89% identity in 397 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein in frr 3'region"
/protein_id="CAD08678.1"
/db_xref="GI:16501501"
/translation="MKQLTILSTGSGTCLDVBHHNPDSPFRVIALVAGKNVARMAE
QCLEFSPYAVMDTSSAEOLKIMLQHGSRTEVLGQQAACMAALDEVGHVMAAIV
GAAGLLPTLAAIRAGTKTILLANKESLVTGCRLEFMDVKRSNARLLPVDSEHNAIFQSL
POSIOHNLGYADLEQNGVTSILLTGSGGPRETPMCDLAAMTPDOACRHPNWSMKRKI
SVDSATMMNKGLEYIEARWLFNASKQMEVLIHPQSVIHSNRYQDSVLAQLGEPDM
RTPIAHTMAMPNRYTGAQPLDFCKLSALTSFAPDYQRYPCLLKAMEAFEGQQAATTA
LNAANEITVAEFLAQQIRFTDIAGNLAVLERMDLHEPASVDDVLDVDAIAREVARKO
VIRLSR"
4064..4087
/notes="dxr"
/notes="PS00017 ATP/GTP-binding site motif A (P-loop)"
5240..5998
/notes="STY0244"
5240..5998
/notes="STY0244"
/notes="Orthologue of E. coli UPPS_ECOLI; Fasta hit to UPPS_ECOLI (253 aa), 94% identity in 252 aa overlap"
/codon_start=1
/transl_table=11
/product="undecaprenyl pyrophosphate synthetase"
/protein_id="CAD08679.1"
/db_xref="GI:16501502"
/translation="MLSATQPVSENLPAHGCRHVAVIMDNGRWAKQKQIRAPGHKA
GAKSVRAVSFAANGIDALTLYAFSENWNRPAQEVSAIMELFWALDSEVSKSLHRL
NVRLRIGDISREFNSRLQERIRKSEALTHTGLTLNIAANYGGWMDIVQGVQLAEL
VOAGVLRPQDIDERLQOQICMHELAPVDLIVIRTGGEHRISNFIWMQIAYAELYFTDV
LWPDFDQDFEGALHAFANRERFEGTGPCDDKA"
5306..5971
/notes="STY0244"
/notes="Pfam match to entry PF01255 UPP_synthetase, Putative undecaprenyl diphosphate synthase, score 492.70, E-value 2.8e-144"
5804..5857
/notes="STY0244"
/notes="PS01066 Undecaprenyl pyrophosphate synthetase family signature"
6011..6868
/notes="STY0245"
/notes="cdsA"
6011..6868
/notes="STY0245"
/notes="Fasta hit to YNBB_ECOLI (298 aa), 33% identity in 298 aa overlap"
Orthologue of E. coli cdsA (CDSA_ECOLI); Fasta hit to

```

alignment\_scores:      Quality: 10.00      Length: 10  
                             Ratio: 1.000      Gaps: 0  
                             Percent Similarity: 100.000      Percent Identity: 100.000

alignment\_block:  
US-09-701-711-2 x AL627266      ..  
Align seg 1/1 to: AL627266 from: 1 to: 268050

648 ProPheTyRGluAsnPhetYrAlaGlyGly 657  
|||||  
10187 CCGTTCATGAGACTTCTATGCCCGTGGT 10216

seq\_name: gb\_ba:AP002550



```

seq_documentation_block:
LOCUS      AP002550      281530 bp      DNA      linear      BCT 07-MAR-2001
DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 1/20.
ACCESSION  AP002550 BA000007
VERSION    AP002550.1 GI:13359456
KEYWORDS
SOURCE     Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RMD 0509952)
           DNA.
ORGANISM   Escherichia coli O157:H7
           Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
           Escherichia.
REFERENCE  1 (sites)
AUTHORS   Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
           Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,T., Abe,H., Iida,T.,
           Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
           Sasaki,C. and Shinagawa,H.
TITLE     Complete nucleotide sequence of the prophage vM2-Sakai carrying the
           verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
           derived from the Sakai outbreak
JOURNAL   Genes Genet. Syst. 74 (5), 227-239 (1999),
MEDLINE   20198780
REFERENCE  2 (sites)
AUTHORS   Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
           Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
           Hayashi,T.
TITLE     Comparative analysis of the whole set of rRNA operons between an
           enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
           Escherichia coli K-12 strain MG1655
JOURNAL   Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
MEDLINE   20557356
REFERENCE  3 (sites)
AUTHORS   Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
           Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
           Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasaki,C. and
           Shinagawa,H.
TITLE     Complete nucleotide sequence of the prophage vM1-Sakai carrying the
           Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
           O157:H7 strain derived from the Sakai outbreak
JOURNAL   Gene 258 (1-2), 127-139 (2000)
MEDLINE   20564182
REFERENCE  4 (sites)
AUTHORS   Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
           Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
           Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasaki,C.,
           Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
           Shinagawa,H.
TITLE     Complete genome sequence of enterohemorrhagic Escherichia coli
           O157:H7 and genomic comparison with a laboratory strain K-12
JOURNAL   DNA Res. 8 (1), 11-22 (2001)
MEDLINE   21156231
REFERENCE  5 (bases 1 to 281530)
AUTHORS   Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
           Hayashi,T.
TITLE     Direct Submission
JOURNAL   Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
           Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
           Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
           URL:http://www.gen-info.osaka-u.ac.jp/,
           Fax:81-6-6879-2047)
COMMENT   genome project.
FEATURES   Location/Qualifiers
           1..281530
           /organism="Escherichia coli O157:H7"
           /strain="O157:H7"
           /sub_strain="RMD 0509952"
           /db_xref="taxon:83334"
           190..273
           /gene="Ecs0001"
           190..273
           /gene="Ecs0001"
           /note="similar to THRL_ECOLI g11786182 percent similarity
           100 in 21 aa, but has 6 additional residues (Conserved in
           E.coli K-12)"

```

gene  
CDS

gene  
CDS

gene  
CDS

gene

```

/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="thr operon leader peptide"
/protein_id="BAB33424.1"
/db_xref="GI:13359457"
/translation="MKRISTTTTITTTTITTTITTTITTTGAG"
354..2816
/gene="Ecs0002"
354..2816
/gene="Ecs0002"
/note="similar to THRA_ECOLI g11786183 percent identity
99 in 820 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="aspartokinase I-homoserine dehydrogenase I"
/protein_id="BAB33425.1"
/db_xref="GI:13359458"
/translation="MRVUKFGTGSVANAERFLRVADILESNAHQGVATVLSAPAKIT
NHLVAMIEKTIISQDAPLNIISDAERIFAELFLGAAQPGFLAQILKTEVDFEFAQIK
HVLHGISLILGQCPDSINAALICRGEKMSIIMAGVLEARGHNVTWIDPEKLLAVGHY
LESTVDIAESTRIIAASRIPADHVMVLMAGFTAGNEKGLVILGRNSDYSAVLAACL
RADCEIWTDDVGVTCOPROVPDARLLKMSYQOAMLSYFGAKVLHPRTTPTIAQF
QIPCLIKTGNPQAFGLIGASRDEDELVKGISLNNMAMFSVSGPCMKGMVMAAR
VFAAMSRARISVVLITQSSEYSISFCVPQSDCVRAERAMQEEFYLELKEGLEPLAV
TERLAIISVVGDMRTLRGISAKFFAALARANINIVIAOGSSERSISVYVNNDDATT
GVRTVHQLMFDNTDQVIEVIGVGGVALLBQLKQKQSWLNKHKHIDLRVGVANSKA
LLTNVHMLNENQOEELAQAPFNRLIRLVKQVHLLNPVTDCTSSQAVADYAD
FLREGFHVTPNKKANTSDMDYHLLRHAARESRKFLYDYNVAGLPIENLONLLN
AGDELMKFSGLISLSIFGLDGMSEATTLAREMGYTEPPDRDLSDMDVARK
LLILARETGRELELADIEIEPVLPAEFNAEGDVAAFMNLSQLDDLLFAARVAKARDEG
KVLRYGVNIDEGVCVRIAEDGNDPLFKVKNGENALAFYSHYIOPYLPLVLRGYGAG
NDVTAAGVFADLLRTLTSWKLGV"
2818..3750
/gene="Ecs0003"
2818..3750
/gene="Ecs0003"
/note="similar to THRE_ECOLI g11786184 percent identity
99 in 310 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="homoserine kinase"
/protein_id="BAB33426.1"
/db_xref="GI:13359459"
/translation="MVKVYAPASSAMSVSDFDLGAATVPDGLIGDVVTVESAEFTF
SLNNLGRFADKLUPSEPRINIVYQCWFERCOELGKQIPYAMTLKKNPIGSGLGSSACS
VVAALMANNEHCCKPINDTRLILALMELEGRISGTHYDNVAPCTGLGGMLTEENDI
VSOQVPGFDEWLWVLAIPGIVKYSTAEARAILPAQYRRQDCIAHRLHAGFIHACYSRQ
PELAARKMDVIAEPIREKLPGFRQARQAAVEIGAIVASGISGSGPTLFCALCKPDPA
QRVADWLKKNYLQNGEGFVHICRLDTAGARVLEN"
3751..5037
/gene="Ecs0004"
3751..5037
/gene="Ecs0004"
/note="similar to THRC_ECOLI g11786185 percent identity
99 in 428 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="threonine synthase"
/protein_id="BAB33427.1"
/db_xref="GI:13359460"
/translation="MKLYNLKDNHNEQVSFAQAVTOGLGKNQGLFFPHDLPEFSLTEID
EMLKDFVTRSARKILSARFIDEIPQELFEERVRAAFAPAPVANYESDVGCLELFGP
TLAFKDFGGRFMAQMLTHIAGDKPVTILLTATSGDTGAAYAHAFYGLPNKVILYPRG
KISPLQEKLFCTLGNIEITVADGDFDAQALVQAQFDDDEELKVALGLNANSINLSR
LQAQICYFEAVQAQLPQARNQLVSVSPNGFDLTAGLSKLGLPLVKRFTAAATNVN
LTVPRFLHDGOWSPKATQATLSNMDVSNQNNPVEELFRKIKWLKELGAAVADDE
TQQQPMRELERKELGYTSEPHAAVAYRALRDLNPGYGLFGLTGAHPAKFESVEAILGE
TLDLPKELAEARADLPLLSHNLPAFDAALRKLMMNHQ"
5251..5547

```



CDS	/gene="ECs0005" 5251..5547 /gene="ECs0005" /note="similar to B0005_ECOLI gi 1786186 percent identity 95 in 98 aa (Conserved in E.coli K-12)" /codon_start=1 /transl_table=11 /evidence=not experimental /product="hypothetical protein" /protein_id="BAB33428.1" /db_xref="GI:13359461" /translation="MKKMSIVLALSILVLVAPMATQAAEITLVPVKLQIGDRNRGY YWDGHRDHGMWKQYEWGRNHPHPGPPPPRHHKKAHHDHGGHGGPKGKHH" /gene="ECs0006" complement(5700..6476) /gene="ECs0006" complement(5700..6476) /note="similar to YAAJ_ECOLI gi 1786187 percent identity 99 in 258 aa (Conserved in E.coli K-12)" /codon_start=1 /transl_table=11 /evidence=not experimental /product="hypothetical protein" /protein_id="BAB33429.1" /db_xref="GI:13359462" /translation="MLILSPAKTLDYQSPLTTRTYTLPELDNSQOLIHEARKLTTP QISTLMRISDKLAGINAAHFDMQDPDTPENARQATILAKGDVYTGLOAETFESEDDFD FAQOHLRMLSGLYGLRPLDLMQPYBLEGIRLENARGKDLQFWDGDIITNKLNEALA AGDNNVYNLASDEYKSKPKLNAEIKPFLVDENKRGFKIISFYAKKARGLMSRF IIGNTKPEQUTGSEGYFFDEASSNGELVFKRYEQR" /gene="ECs0007" complement(6546..7976) /gene="ECs0007" complement(6546..7976) /note="similar to YAAJ_ECOLI gi 1786188 percent identity 99 in 476 aa (Conserved in E.coli K-12)" /codon_start=1 /transl_table=11 /evidence=not experimental /product="putative inner membrane transport protein" /protein_id="BAB33430.1" /db_xref="GI:13359463" /translation="MPDFESFINSLWGSVMYLLFGAGCWFRTFTGFGVQFYIROFG KSLKSIHPQPGGLTSFQSLCTSLAARVSGNLGVALAITAGPGGAVFWMVAAFGI MATSFACSLAOLYKERDVNGFRGGPANTMARGLMRWGVLFAVELLIAYGLIFSG VOANAVARALSFSEDPPLVGTGILLAVFALLIICGLHGVARLMOGFVPLMAIIVLT SLVICVINIGOLPHVTSIFESAFQOEAGGAGTYLSQATNGFQRFMSNEAGMG STPNAAAAASPPHPAAGIVOMIGIFIDTIVICTASAMLILLAGNGTTPLEGIQ LLOKARVLMGSGAEFVTLVILFAFSSIVANYIYAENLFFLRNNPKAIWCLRIC TPTATVIGGTLISLPLMWLADITMACMAITNTAILLSPPVHTTASDYLRRKLGVR PVFDLPYPDIGRLSRDWDVDSQ" 8255..9208 /gene="ECs0008" 8255..9208 /gene="ECs0008"	LOCUS DEFINITION ACCESSION VERSION KEYWORDS SOURCE ORGANISM REFERENCE AUTHORS FEATURES source mRNA gene	AE003597 Drosophila melanogaster genomic scaffold 14200013386036 section 7 of 9, complete sequence. AE003597 AE002647 AE003597.1 GI:7296525 HTG. fruit fly. Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 295566) Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F., George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., Brandon R.C., Rogers Y.H., Blazer R.G., Champe M., Pfeiffer B.D., Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor Miklos G.L., Abril J.F., Agbayani A., An H.J., Andrews-Pfannkuch C., Baldwin D., Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S., Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P., Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferrieri S., Fleischmann W., Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C., Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M., Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H., Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissbach J., Williams S.M., Woodge T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Ye R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M. and Venter J.C. The genome sequence of Drosophila melanogaster Science 287 (5461), 2185-2195 (2000) 20196006 2 (bases 1 to 295566) Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M. and Venter, C.J. Direct Submission Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA Location/Qualifiers 1..295566 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /chromosome="3L" join(<27808..27909,28034..28318,28822..>29085) /gene="CG12673" /product="CT35415" /db_xref="FLYBASE:FBan0012673" /db_xref="FLYBASE:FBan0037153" /evidence=not experimental <27808..>29085 /gene="CG12673" /map="79D4-79E1" /db_xref="FLYBASE:FBan0012673"
-----	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

```
/db_xref="FLYBASE:FBgn0037153"
/evidence-not_experimental
join(27808..27909,28034..28318,28822..29085)
/gene="CG12673"
/note="CG12673 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0012673"
/db_xref="FLYBASE:FBgn0037153"
/evidence-not_experimental
/protein_id="AAF51810.1"
/db_xref="GI:7296526"
/translation="MGRVGRRAFGRGAPPAQRCIRVHVSNNDIIAANDTMRLLYMY
HAQPPHGSVRPGLPDPARAFRPRMVLQRAQLPMPSTHDERVRVLELRNEDVE
LPAPDLFLKMFKEIDINKRHLRIEPIYDSSSVHYLQHTLHECOGAHALEEE
MARQGPCLGARGIPIACNAIVASWSRGSEVTSIRERNONLPPTYPCQNSCPQT"
join(<35333..35500,35594..35756,35909..35937)
/gene="CG7495"
/product="CT23023"
/db_xref="FLYBASE:FBan0007495"
/db_xref="FLYBASE:FBgn0037154"
<35333..35500,35594..35756,35909..35937)
/gene="CG7495"
/map="79E1-79E1"
/db_xref="FLYBASE:FBan0007495"
/db_xref="FLYBASE:FBgn0037154"
join(35333..35500,35594..35756,35909..35937)
/gene="CG7495"
/note="CG7495 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0007495"
/db_xref="FLYBASE:FBgn0037154"
/protein_id="AAF51811.1"
/db_xref="GI:7296527"
/translation="MDPNWRHIIPPGOKRVVSEQCIEDCTGYAFPOGGINIAFMMVR
THQIGKEVKLRQIQTELPPIADHSNIDVAYQDFRLPQSVHSMGPDRLIAECIYDS
SSRAITIGDEVPALES"
join(<48427..48474,48682..48989,49294..49318)
/gene="CG14461"
/product="CT34144"
/db_xref="FLYBASE:FBan0014461"
/db_xref="FLYBASE:FBgn0037155"
/evidence-not_experimental
<48427..49318)
/gene="CG14461"
/map="79E1-79E1"
/db_xref="FLYBASE:FBan0014461"
/db_xref="FLYBASE:FBgn0037155"
/evidence-not_experimental
join(48427..48474,48682..48989,49294..49318)
/gene="CG14461"
/note="CG14461 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0014461"
/db_xref="FLYBASE:FBgn0037155"
/evidence-not_experimental
/protein_id="AAF51812.1"
/db_xref="GI:7296528"
/translation="MWAEISDGTTLCSLSECEYSINVTKYKKHRCKPKRPL
APPTERTAPPASDLSLPVLHEDLNNNIEGARSSSATDVHLSRSGSRHFISC
LLWLGSWSWLLMLRTPPTPYTC"
complement(join(64112..66736,66987..69059,69117..69239))
/gene="BcdNA:GH03694"
/product="CT26020"
/db_xref="FLYBASE:FBan0009063"
/db_xref="FLYBASE:FBgn0028500"
complement(<64112..69239)
/gene="BcdNA:GH03694"
/note="CG9063"
/map="79E1-79E1"
/db_xref="FLYBASE:FBan0009063"
/db_xref="FLYBASE:FBgn0028500"
complement(join(64502..66736,66987..69041))
/gene="BcdNA:GH03694"
```

```
/note="BcdNA:GH03694 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0009063"
/db_xref="FLYBASE:FBgn0028500"
/protein_id="AAF51813.1"
/db_xref="GI:7296529"
/translation="MYTPVGPWKRVGLALPGESASIRIHCCDAVKILVAAYGDDFLGI
WYANPLPIAYFRFTEDSROYGANOLIVKRPDSRQALTLASGLLQLDLEFANGM
GILQIDPPAASLRKDAELFIKENIPRLSLRCLSVTLGSVITTVCCISLSEILLAT
QSCELLRLQWTELEHAENDLELPALSSIKLRDIPFYVOOQOQARNVPLNRSYVA
SLEYSPITGGCAAVSDRRAFLJANHILRFETHMHGFMWPDVEDASVCSVNHKFRLL
AYGESAVKVIADDAITGGLFESHRLITENILPDSLSVNEKLSKSPDGCYLAWSWT
NGUSLWSTFGALLMSTLSDWFDNVDVLCQNPILKIRLEWSTEGYOLFMLKLHPKED
KSNVLOQFVKSALSMNPMCTTSPHILLQDDCLYLNOGNLEITYAGSHGFTFSSGL
GSDSDISGDCLSELKQSPHTGSILTSTESKWTVLQPLNYAATNWPRIYAAIDPDGLH
LAVAGRTGLAHSVLTWRWKLFGNESOEKDFVSGGLLWHGCVVMGYSLLDRTDEL
RCYPADCKLDNOYGHKLQVRAPVLSLASFHOLIVLTADGIVSLFNMKSNAYALDIE
CAYELDVKSCIHACIVSLVTNLKNELKFPQGLGGDAETIIVNCGRILMQORDA
GEQVPNTLLATCLASCEVEFWLSHSLERCAMRDCWLISYGAHMRVLPILPPGRERR
EGEQGAQLHSFNSKRIIMSLFPLKPLVPLVDFNVIYLVENESTLYANQVSHFSL
PFAYMRKSQIYLHKVRLQIKRNLGYSANEMAOCCSLPYFPHALELLLEVEEEA
TSKOPIPDAQLPSILDFIRFPVPLETVQCARKEITALWPLYLFSMACKPKDLQMCCL
QSEQLDTAASVLIILIONLEPSVSKQYATMLLDIALQOKWELAKDLIRFLKADPNE
IDSPSSWVNVKIAPPVNTQQVQONDAFNMVLGPIARERSFTVTSNLPKOK
QASPTPGVAPVTSSAGAPSVRRRTKQRETCIDILQIQRHARQLLQNHKMLDLGY
MCAYLDLHLSWLSQESERAAKDDFAGALQALHEELDLPFPFPAKDDPAQIRGSL
ROTCGGSSOTSSECYFSLATPNGAATQSPOLQPSIRREELELOQSPSLPLKTRSGS
QLSFDNFRYRLYSLPTSEDDLAVDILPQKLSIKRLVLLQLFIBANCTDYALVLSILL
QDAASISRVNGIIRSESVHTRCTESALKQLSQTSEHSGSLYRGFVLTLOPHVYLL
EQYIOSIGDAPCSQLQDAGPTGQGVDSVTLQGLQNEVEGFVPSQOANQMTVAD
LNPQHRLRTHASLESNGNAVASGSAHSPTPQRLSRONSREREGCRM"
join(69753..69861,69916..69967,70021..71690,71763..72208)
/gene="Ddx1"
/note="Nucleotide sequence of the Cclera sequence differs
from the published sequence for this transcript."
/product="CT25986"
/db_xref="FLYBASE:FBan0009054"
/db_xref="FLYBASE:FBgn0015075"
<69753..72208)
/gene="Ddx1"
/note="CG9054"
/map="79E1-79E2"
/db_xref="FLYBASE:FBan0009054"
/db_xref="FLYBASE:FBgn0015075"
join(69846..69861,69916..69967,70021..71690,71763..72208)
/gene="Ddx1"
/note="Ddx1 gene product; Nucleotide sequence of the
Cclera sequence differs from the published sequence for
this transcript"
/codon_start=1
/db_xref="FLYBASE:FBan0009054"
/db_xref="FLYBASE:FBgn0015075"
/protein_id="AAF51814.1"
/db_xref="GI:7296530"
```

## alignment\_scores:

Quality:	10.00	Length:	10
Ratio:	1.000	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

## alignment\_block:

US-09-701-711-2 x AE003597 ..

Align seg 1/1 to: AE003597 from: 1 to: 295566

789 SerTyrAlaLysProLeuAsnLysLysGln 798

|||||

162573 TCTTATGCAAGCCACTTAACAAAAGCAA 162602

seq\_name: gb\_pat:AX078597

seq\_documentation\_block:

LOCUS AX078597

2019 bp DNA linear PAT 22-FEB-2001

DEFINITION Sequence 111 from Patent WO0107624.  
ACCESSION AX078597  
VERSION AX078597.1 GI:13158239  
KEYWORDS  
SOURCE Pseudomonas putida.  
ORGANISM Pseudomonas putida  
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
Pseudomonas.  
REFERENCE 1 (bases 1 to 2019)  
AUTHORS Fraser C.M., Venter C., Tuemmler B., Hoheisel J., Duesterhoeft A., Hilbert H., Timmis K.N., Moore E., Straetz M., Helm S., Nelson K.E., Hickey E. and Peterson J.  
TITLE Dna sequences which are suited for specifically detecting Pseudomonas putida kt2440  
JOURNAL Patent: WO 0107624-A 111 01-FEB-2001;  
THE INSTITUTE FOR GENOMIC RESEARCH (US); QIAGEN GmbH (DE); Gesellschaft fuer Biotechnologische Forschung mbH (GBF) (DE); Deutsches Krebsforschungszentrum (DKFZ) (DE); Medizinische Hochschule Hannover (DE)  
FEATURES  
source  
1. .2019  
Location/Qualifiers  
/organism="Pseudomonas putida"  
/db\_xref="taxon:303"  
BASE COUNT 404 a 671 c 603 g 338 t 3 others  
ORIGIN  
  
alignment\_scores:  
Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-09-701-711-2 x AX078597 ..  
Align seg 1/1 to: AX078597 from: 1 to: 2019  
368 AspGluValLeuArgArgGluMetArg 376  
|||||  
250 GACGAAGTGTGCTGCGTCGCGAAATGCGC 276  
seq\_name: gb\_pr:AY062939  
seq\_documentation\_block:  
LOCUS AY062939 2418 bp DNA linear PRI 26-DEC-2001  
DEFINITION Homo sapiens RETSDR2 (RETSDR2) gene, promoter region and partial cds.  
ACCESSION AY062939  
VERSION AY062939.1 GI:17978577  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 2418)  
AUTHORS Obeyesekere V.R., Saffery R.E. and Krozowski Z.S.  
TITLE 17#HSDXI is a novel human microsomal dehydrogenase expressed in steroidogenic cells  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2418)  
AUTHORS Obeyesekere V.R., Saffery R.E. and Krozowski Z.S.  
TITLE Direct Submission  
JOURNAL Submitted (15-NOV-2001) Molecular Hypertension, Baker Institute, Commercial Rd, Melbourne, Vic 8008, Australia  
FEATURES  
source  
1. .2418  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
gene <1. >2418  
misc\_feature <1. .2415  
/gene="RETSDR2"  
/note="contains promoter region and 5' untranslated

region"  
region"  
<2416. .>2418  
/gene="RETSDR2"  
/product="RETSDR2"  
2416. .>2418  
/gene="RETSDR2"  
/codon\_start=1  
/product="RETSDR2"  
/protein\_id="AAL48315.1"  
/db\_xref="GI:17978578"  
/translation="M"  
BASE COUNT 731 a 514 c 494 g 678 t 1 others  
ORIGIN  
  
alignment\_scores:  
Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-09-701-711-2 x AY062939/rev ..  
Align seg 1/1 to reverse of: AY062939 from: 1 to: 2418  
46 IleGluSerLeuGlnSerValLeuPro 54  
|||||  
842 ATAGAAAGTTTACAAAGTGTCTTCCA 816  
seq\_name: gb\_pl:ZMZAG2  
seq\_documentation\_block:  
LOCUS ZMZAG2 7071 bp DNA linear PLN 29-SEP-1995  
DEFINITION Z.mays zag2 gene.  
ACCESSION X80206  
VERSION X80206.1 GI:940880  
KEYWORDS MADS-box protein; ZAG2 gene.  
SOURCE Zea mays.  
ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE 1 (bases 1 to 7071)  
AUTHORS Theissen G., Strater T., Fischer A. and Saedler H.  
TITLE Structural characterization, chromosomal localization and phylogenetic evaluation of two pairs of AGAMOUS-like MADS-box genes from maize  
JOURNAL Gene 156 (2), 155-166 (1995)  
MEDLINE 95278740  
REFERENCE 2 (bases 1 to 7071)  
AUTHORS Theissen G.  
TITLE Direct Submission  
JOURNAL Submitted (22-AUG-1994) G. Theissen, Max Planck Institut fuer Zuechtungsforschung, Carl von Linne Weg 10, D-50829 Koeln, Germany  
REMARK Revised by [3]  
REFERENCE 3 (bases 1 to 7071)  
AUTHORS Theissen G.  
TITLE Direct Submission  
JOURNAL Submitted (11-AUG-1995) G. Theissen, Max Planck Institut fuer Zuechtungsforschung, Carl von Linne Weg 10, D-50829 Koeln, Germany  
COMMENT On Aug 13, 1995 this sequence version replaced gi:854643. Sequences overlapping with: R. Schmidt; Plant Cell, Vol. 5, 729-737, 1993.  
FEATURES  
source  
1. .7071  
Location/Qualifiers  
/organism="Zea mays"  
/sub\_species="mays"  
/db\_xref="taxon:4577"  
/chromosome="3"  
/map="66"  
/clone="EMBL4-II7b"  
/clone\_lib="EMBL4"  
GC\_signal 168. .177

```
GC_signal      /note="putative"
295. .304
GC_signal      /note="putative"
311. .320
TATA_signal    /note="putative"
392. .399
TATA_signal    /note="putative"
917. .924
TATA_signal    /note="putative"
1288. 1297
TATA_signal    /note="putative"
2058. .2064
/note="putative"
Join(2133. .2314,4276. .4360,4755. .4819,4959. .5058,
5281. .5322,5411. .5452,5610. .5900)
/gene="ZAG2"
2133. .2314
/gene="ZAG2"
/number=1
Join(2133. .2314,4276. .4360,4755. .4819,4959. .5058,
5281. .5322,5411. .5452,5610. .5900)
/gene="ZAG2"
/codon_start=1
/protein_id="CAA56504.1"
/db_xref="GI:1001934"
/db_xref="SPTREMBL:Q43353"
/translation="MGRGRIETKRIENNTSRQVTFCKRRNGLLKAYELSVLCDAEVA
LIVFSRRLVEYANVSQVATVRYKKAHTVGSSSGPPLLHNAQOQYQOFSAKLRNQ
IOMLQNTNRHLVGVDSVGNLSKELKQLSRLKLGSKTRAKSELLAAEISYMAKRET
BLQNDHWTLRKIERGEQLOQVTVARVAAAATNLEAPFLFLEMDTKCFCTGGPFA
TIDMKCFUFGSLQOMLEAQOQMLATELNILGYQLAPPDGAANNPHQF"
4276. .4360
/gene="ZAG2"
/number=2
4755. .4819
/gene="ZAG2"
/number=3
4959. .5058
/gene="ZAG2"
/number=4
5281. .5322
/gene="ZAG2"
/number=5
5411. .5452
/gene="ZAG2"
/number=6
5610. .5900
/gene="ZAG2"
/number=7
5898. .6197
polyA_signal   1813 a 1583 c 1526 g 2144 t 5 others
BASE COUNT    6060. .6066
ORIGIN

alignment_scores:
  Quality: 9.00      Length: 9
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x ZMZAG2/rev ..
Align seg 1/1 to reverse of: ZMZAG2 from: 1 to: 7071
      489 ValSerGlnSerLeuSerGlyTyrTyr 497
      |||||
2913 GTCAAGTCAATCTTCTCTGGGTATAT 2887

seq_name: gb_pl:SPCC188
seq_documentation_block:
LOCUS          SPCC188              35412 bp    DNA    linear    PLN 19-JAN-2000
```

DEFINITION  
ACCESSION  
VERSION  
KEYWORDS

S.pombe chromosome III cosmid cl188.

AL049662

AL049662.1 GI:4678680

5s rRNA: alpha-amylose; GTP cdc domain: LTR; nuclear protein;  
phosphatase; pseudogene; ribonuclease; septin homolog; signal  
recognition particle; srp54; tRNA; ubiquitin carboxyl-terminal  
hydrolase.

## SOURCE

fission yeast.

## ORGANISM

Schizosaccharomyces pombe

Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

Schizosaccharomycetales; Schizosaccharomycetaceae;

Schizosaccharomycetes.

1 (bases 1 to 35412)

Lyne,M., Rajandream,M.A., Barrell,B.G., Seeger,K. and Harris,D.

Direct Submission

Submitted (16-APR-1999) European Schizosaccharomycetes genome

sequencing project, Sanger Centre, The Wellcome Trust Genome

Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk

## NOTES

Details of yeast sequencing at the Sanger Centre are available on

the World Wide Web.

(URL, http://www.sanger.ac.uk/Projects/S.pombe/)

During 1995 to 1996 about 66% of S. pombe chromosome I was

sequenced by the Sanger Centre. The sequencing of the S. pombe

genome is now being continued with funding from The European

Commission. Fourteen European sequencing laboratories, including

the Sanger Centre, are participating in the project.

Protein coding regions (CDS) have been predicted with the help of

computer analysis using the Genefinder program in PomBase (an ACEDB

database) with additional predictions for the branch-acceptor sites

supplied by the program Splice. CAUTION: It is possible that for

any individual CDS we may have underestimated or overestimated the

number of introns/exons or we may not have chosen the correct

splice donor/acceptor sites.

CDS are numbered using the following system eg SPBC25H2.01c. SP (S.

pombe), B (chromosome 2), c25H2 (cosmid name), .01 (first CDS), c

(complementary strand).

The more significant matches with motifs in the PROSITE database

are also included but some of these may be fortuitous.

The length in codons is given for each CDS.

IMPORTANT: This sequence MAY NOT be the entire insert of the

sequenced clone. It may be shorter because we only sequence

overlapping sections once, or longer, because we arrange for a

small overlap between neighbouring submissions.

## FEATURES

## source

1. .35412

/organism="Schizosaccharomyces pombe"

/strain="972h"

/db\_xref="taxon:4896"

/chromosome="III"

/map="IIIR"

/clone="cosmid cl188"

complement(1. .610)

/gene="SPCC188.01c"

/note="SPCC11E10.09c"

1. .2559

## misc\_feature

/note="nominal overlap with cosmid SPCC11E10, EM:AL121783

S. pombe chromosome 3"

complement(1. .610)

/partial

/gene="SPCC188.01c"

/note="SPCC188.01c, len&gt;202,

SIMILARITY:Schizosaccharomyces pombe, O13996, putative

alpha-amylose precursor, (491 aa), fasta scores: Opt: 833,

E():0, (60.9% identity in 184 aa)"

/codon\_start=1

/label="SPCC188.01c

/product="putative alpha-amylose precursor"

/protein\_id="CAB41221.1"

/db\_xref="GI:4678681"

/translation="MNLICSLIPKKPSYKLWRKQVIYQVLTDRFALDEDFYAKAS

GNLYLGGTWKGTIRNLDIYKISGCTAIWISPIVKNISITDGGQAVHGVWADMTOLN

ENFGTEEDLKELVNAIHEKNMLCMVDIVVNMHGAGSKPVPFLQLQPFNSGYHNWQ

```

/rRNA
complement(1533..1696)
/notes="S. pombe 5s rRNA gene and flanks"
/rRNA
2318..2400
/notes="trna Lys anticodon CTT"
/product="trna Lys"
join(3061..3084,3128..3223,3264..3477,3809..4563,
4616..4914,5004..5262)
/genes="SPCC188.02"
CDS
join(3061..3084,3128..3223,3264..3477,3809..4563,
4616..4914,5004..5262)
/genes="SPCC188.02"
/notes="SPCC188.02, len:547, SIMILARITY:Oryctolagus
cuniculus, Q28653, protein phosphatase pp2a0 b' subunit
gamma isoform., (586 aa), fasta scores: opt: 1779, E():0,
(52.3% identity in 507 aa)"
/codon_start=1
/label="SPCC188.02"
/product="putative protein phosphatase subunit"
/protein_id="CAB41222.1"
/db_xref="GI:4678682"
/translation="MKGIKSKMLSRGKSQDTQKSKKSKSKNSHDSKAPKESPT
DNGSVIGAQNDPLVPHKSGKVPIDTPTPRDEITLLENVTRVKORSLSYHSENH
NLVRLPSTFDVNVKWHSLALEKLEQCCVDFNPDPSTLYCKEVRKALQDLIDLIS
YKKEALDESILYSIVHMAVNVFRPLPPSPNPPEIMDEDEPALEVAWPHLHLYD
FPLRFESPLATSAKVYINQKIFIKLLVLFDSDEPRDFLTKTLHRIYKFLSLR
AFIRRSINLFTQYVYENQFNGIAELLEILGSIINGFALPKKEHKIFLSVGLIPLH
KAKSLPYQIAYGVQVQVEWGLLRYWPKVNSSEVLFLNLEDIEIIE
VMEPSEFLKIQVPLHKLATSISSONFOAERALLYFNNDYFVHLVEENVDLIPIY
PALFELSKSHNRVYHSMVCNVILKFMIDNPLSFLFDEVDAYSESRKKEDEBILREER
WTLLENIAKENAMKLSQNPTTVHSTTERLKLKSLDYTNG"
3085..3090
/misc_feature
/genes="SPCC188.02"
/notes="gtaaga, splice donor sequence"
3113..3127
/misc_feature
/genes="SPCC188.02"
/notes="ttaaacaagaatctag, splice branch and acceptor"
3224..3229
/misc_feature
/genes="SPCC188.02"
/notes="gtacagg, splice donor sequence"
3249..3263
/misc_feature
/genes="SPCC188.02"
/notes="ttaaattttcatag, splice branch and acceptor"
3478..3483
/misc_feature
/genes="SPCC188.02"
/notes="gtatgt, splice donor sequence"
3792..3808
/misc_feature
/genes="SPCC188.02"
/notes="ctaacctgtctctatag, splice branch and acceptor"
4564..4569
/misc_feature
/genes="SPCC188.02"
/notes="gtatgt, splice donor sequence"
4598..4615
/misc_feature
/genes="SPCC188.02"
/notes="ctaacagtgaacaatag, splice branch and acceptor"
4915..4920
/misc_feature
/genes="SPCC188.02"
/notes="gtaagt, splice donor sequence"
4989..5003
/misc_feature
/genes="SPCC188.02"
/notes="ctgacgtttgttag, splice branch and acceptor"
join(5588..7501,7542..7673,7734..7962,8011..8120,
8183..8425)
CDS
join(5588..7501,7542..7673,7734..7962,8011..8120,
8183..8425)
/genes="SPCC188.03"
/notes="SPCC188.03, len:874, SIMILARITY:Saccharomyces
cerevisiae, Q06680, chromosome iv cosmid 9798., (1051 aa),
fasta scores: opt: 406, E():2.4e-18, (30.0% identity in
949 aa)"
/codon_start=1
/label="SPCC188.03"

```

```

/product="hypothetical protein"
/protein_id="CAB41223.1"
/db_xref="GI:4678683"
/translation="MSCIQLISSOTSIAHRKLCNKLETLTFOEGFETDLRALNII
LVKGVNSADRVLRFLVTFVNYLQKDEIDIVQILKHLGLDARKDQTVYRCQO
IARVNVCKEIDDDLYNTLTKELLSRVLDRESIVLEAVVLSLQEDTGDSENDVR
NLLFLQNDSPSEVRVLLNIEVNSNTPLFLERARDVAAARCKVCYARVLPKIGD
FYLISKRRVRIILKGLNDRDESVEKAADMLAYQWENADNLLLELLERLDYVNSND
VAVLAIKFFDVRVDSLSQLEPPEQFWLELTAESSLARTFNEICTEKNTDLDKMP
ENVOLTYTEROVYSLRDKSSYDESCFIIETOLLYIGLSODMVDIEGRKLLKSLTNSL
SMALPDSLISLHIELLRKLCSENDFCSLLVEITTEVEQHSQNTQEQGSNAPE
LNKNDTEGEITVSKSPSPSPNPNEPEPDMDGKYKEAFNELKCLSTVQCCLFENI
TSLNENLYVMVMDLKTLLIPAVSHDLPIREKGLCISLVCLLNADLAFENVPLYLHC
YKGLSVKLTARTITMDLIQHKAKFTEYDAISILFEALGEFENAELOTLGAE
IAGLIVILHYRDELFLKPLTIQVFEPTVDNHALRVGLVFFVYFAFGAHENQWRAT
IFCDALLSLEIYRDLDEDSVOLSIGIAQMDLWDNKLRYKRTQTDGDIYALNHN
VHLHLAMFISLPNASEGKRFMSLLGKLKIPDLPSSDYQRTKRLKETYESHGF
TWDSTSLILAKFERMLLQNEARSKFEETEERLMEAEENEHAGAEASIGELIPT
VEANMEDEBEVYKOEEDL"
7502..7507
/misc_feature
/genes="SPCC188.03"
/notes="gtatgt, splice donor sequence"
7528..7541
/misc_feature
/genes="SPCC188.03"
/notes="cttaacaacaactag, splice branch and acceptor"
7674..7679
/misc_feature
/genes="SPCC188.03"
/notes="gtaagg, splice donor sequence"
7719..7733
/misc_feature
/genes="SPCC188.03"
/notes="ttaaagcagcttttag, splice branch and acceptor"
7963..7968
/misc_feature
/genes="SPCC188.03"
/notes="gtgggg, splice donor sequence"
7995..8010
/misc_feature
/genes="SPCC188.03"
/notes="ctaacacaacaatgtag, splice branch and acceptor"
8121..8126
/misc_feature
/genes="SPCC188.03"
/notes="gtaagt, splice donor sequence"
8169..8182
/misc_feature
/genes="SPCC188.03"
/notes="ctataatgtag, splice branch and acceptor"
complement(join(8422..8564,8629..9042,9091..9250))
/genes="SPCC188.04c"
CDS
complement(join(8422..8564,8629..9042,9091..9250))
/genes="SPCC188.04c"
/notes="SPCC188.04c, len:237, SIMILARITY:Saccharomyces
cerevisiae, Q12234, chromosome xv reading frame orf
yor216c., (484 aa), fasta scores: opt: 163, E():0.023,
(26.8% identity in 194 aa)"
/codon_start=1
/label="SPCC188.04c"

```

## alignment\_scores:

Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-701-711-2 x SPCC188 ..

Align seg 1/1 to: SPCC188 from: 1 to: 35412

237 LeuasnAlaGlyPheValArgPheGlu 245

|||||

27972 TTAACGCGGGCTTTGTAAGATTGAA 27998

seq\_name: gb\_pl:SPAC8A4

seq\_documentation\_block:

LOCUS SPAC8A4

DEFINITION S.pombe chromosome I cosmid c8A4.

DNA

linear

PLN 04-NOV-1995

# ACCESSION VERSION KEYWORDS

Z66569  
Z66569.1 GI:1052533  
Drosophila PRANUT homolog; Drosophila PUFF homologue; permease;  
RNA helicase; sulphite reductase; ubiquitin carboxy-terminal  
hydrolase.

## SOURCE

ORGANISM  
Schizosaccharomyces pombe  
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
Schizosaccharomycetales; Schizosaccharomycetaceae;  
Schizosaccharomycetes

## REFERENCE AUTHORS JOURNAL REFERENCE AUTHORS TITLE JOURNAL

1 (bases 1 to 38141)  
Lye, G. and Churcher, C.M.  
Unpublished  
2 (bases 1 to 38141)  
Barrell, B.G., Rajandream, M.A. and Walsh, S.V.  
Direct Submission  
Submitted (02-NOV-1995) Schizosaccharomycetes pombe chromosome I  
sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge  
CB10 1RQ E-mail: barrell@sanger.ac.uk

## COMMENT

Details of yeast sequencing at the Sanger Centre are available on  
the World Wide Web.  
(URL, <http://www.sanger.ac.uk/yeast/home.html>)

Protein coding regions (CDS) have been predicted with the help of  
computer analysis using the Genefinder program in PomBase (an ACEDB  
database) with additional predictions for the branch-acceptor sites  
supplied by the program Sp3splice. CAUTION: It is possible that for  
any individual CDS we may have underestimated or overestimated the  
number of introns/exons or we may not have chosen the correct  
splice donor/acceptor sites.

CDS are numbered using the following system eg SPAC5H10.01c. SP (S.  
pombe). A (chromosome 1), c5H10 (cosmid name), .01 (first CDS), C  
(complementary strand).

The more significant matches with motifs in the PROSITE database  
are also included but some of these may be fortuitous.

The length in codons is given for each CDS.

IMPORTANT: This sequence MAY NOT be the entire insert of the  
sequenced clone. It may be shorter because we only sequence  
overlapping sections once, or longer, because we arrange for a  
small overlap between neighbouring submissions.

Cosmid c8A4 is not overlapped by other cosmids.

## FEATURES source

Location/Qualifiers  
1..38141  
/organism="Schizosaccharomyces pombe"  
/db\_xref="taxon:4896"  
complement(join(1..2612,2692..2794,2846..2848))  
/gene="SPAC8A4.01c"  
complement(join(1..2612,2692..2794,2846..2848))  
/note="SPAC8A4.01c", partial orf, len: > 906, similar to  
ubiquitin carboxyl-terminal hydrolases family 2; most  
similar to SC9952X Z49212 S. cerevisiae chromosome XIII  
cosmid 9952, gene YM9952.06, (62% identity in 764 base  
fragment overlap); contains PS00972 Ubiquitin  
carboxyl-terminal hydrolases family 2 signature  
/codon\_start=1  
/product="unknown"  
/protein\_id="CAA91511.1"  
/db\_xref="GI:1052534"  
/db\_xref="SWISS-PROT:Q09879"  
/translation="MVGTGTLVDQSKSLINNDLLNEKLKEDFENVSIDVKIHEELR  
RALPDYEEGRFTWIKSWHELDRAVSPQFVAGSRQKITYFPQGLQSGAFTSI  
FLEYIPSEEEKLSNKGCCOFAFVSNRPKPSLSVANSACRFSPEIVDMGFTQFAE  
LKLCLCQAPDVPPIVEDGALLTAYVRLKDPGTGLVHSPFNQDYSKIATGYGLKQ  
GATCYMNSLLQSLIILHAFRIVYQIPDPSQGDSTAYALQRCFYNLQFNNEPVSTY  
ELTKSFQWDSLDSFMQHDVQEFNVRVLDQNLERSMRDTKVENALTNLFVGRMKSYIAT

NVNFESARSEDYWDIQLNVKGMKNLEDSFRSYIQVETLEGDNCYFADTYFGQEAKKGV  
IFESFPITLHLQKRFYDFEDDMKIKINDRYEFPLEDAKAFSPENQSQNCYVFL  
YGVLVHSGDLHNGHYALLKTEKDGPKYKDDTRVTRATLREVLNENYGGDYIMHPPL  
RSPVKLRFMSAYMLLYLRKDKLDELMPVSADEIPEHLKALNP.SIOLAEURRERL  
ESHLTKVQLITPFEYSHEHEDIADFNAYKEETIPQFRKKKEAFSEFIIVAEKL  
GYPOCMRFYVYVRKHNTVRVSPVNEELNTEEMVEKNVWNSQGLIRLYLEITPENE  
LSSSLTHQNTGEWNAFIFVKYFDRKSQISGCTLHVNKSDEIRISICPLLCRAALPK  
NPLNLIYIEIKPQWDFLRLEKTFQSELSTGDIICFPCRP.SALEDDIVNSGFDLAL  
KLYDFLSKNVLVLEPFRIDODSIIIEFEMLLDRRIKYDDLCTELGQKLGIGADHRLT  
TCNPLTYSAGWVVPNDSNITLYEILYSSSEEMPSNVIFYETMDVSQFTWI"  
complement(2032..2076)  
/gene="SPAC8A4.01c"  
/note="PS00972 Ubiquitin carboxyl-terminal hydrolases  
family 2 signature 1"  
complement(2613..2628)  
/gene="SPAC8A4.01c"  
/note="splice branch and acceptor sequence,  
tactaacgggtccatag"  
complement(2686..2691)  
/gene="SPAC8A4.01c"  
/note="splice donor sequence, gtacgt"  
complement(2795..2812)  
/gene="SPAC8A4.01c"  
/note="splice branch and acceptor sequence,  
ctaacttaataatgtttag"  
complement(2840..2845)  
/gene="SPAC8A4.01c"  
/note="splice donor sequence, gtattg"  
complement(3396..5225)  
/gene="SPAC8A4.02c"  
complement(3396..5225)  
/gene="SPAC8A4.02c"  
/note="SPAC8A4.02c, unknown, len: 609, contains 6 copies  
of a 36 aa repeat; highly similar to possible spliced gene  
or pseudogene SPAC8A4.3c,4c,5c"  
/codon\_start=1  
/product="unknown"  
/protein\_id="CAA91512.1"  
/db\_xref="GI:1052535"  
/db\_xref="SWISS-PROT:Q09880"  
/translation="MLFURFFLFTFTSIFTVVWSEPPVSAKFAKRLVLSATSSSSSTC  
PEFTIFSEGFSEYITTIYPSKSGRGTNHSTIYTTINSTVASTVTVPLADGDV  
VKDEPTAGTVITITSGSHEFTTILAEASGTVPGVEVPELAGTVITITIHSGSVEY  
NTLATASTGTVPGTVVEVPAVGTVITITISGSEFTTTLASAGSISGTVIEPTA  
GTVITITISGQEYTTILAEASGTVPGVEVPAVGTVITITISGSEYTTITLVPAS  
GVSQTVVEVPAVGTVITITISGQSAFTTVPASGVSQTVVEVPAVGTVITITVYE  
GSRITSLTASGTVPGTVITISGQSAFTTVPASGVSQTVVEVPAVGTVITITVYE  
CNGERGLQYAVNYNDISSKNQFCATGTVDSSTQPAVFGSSDLQSSPLFTGLS  
SDSVQMTSSINLPGPPDATAMGTSACKVIYQFFRVPVTDITSLDVTNDDV  
FYGMGDKAISGNSNTYDYAYWHATNGQTGIASFSGSLTADTIVPVRFVANGAG  
KGFDFSFVDSGNSNPTSYATATCTESELFPFGQNGGVDN"  
6993..7359  
/note="sequence similar to tfl LTR"  
complement(8437..9024)  
/gene="SPAC8A4.03c"  
complement(8437..9024)  
/note="SPAC8A4.03c"  
/note="Protein sequence is in conflict with the conceptual  
translation: SPAC8A4.03c, unknown, len: 195, highly  
similar to SPAC8A4.02c, possible pseudogene orf as  
similarity extends through upstream stop codon; or  
possibly spliced to upstream orf"  
/codon\_start=1  
/product="unknown"  
/protein\_id="CAA91513.1"  
/db\_xref="GI:1326030"  
/translation="FYASNGTNTVATFKPAYFGSSDLQSSPLETGVLTSPADLPQW  
TSSYDLPAYPNATVMGNTTKVYIYQFFRVPVSGLTALTINDDVFFGWAQSK  
AIGSNKNYDSYAYWHITTKGTGIASFSGSLTAGSIPIRFVANGAGKGGDFDF  
IGDGTYSNPTSYATCTQIFLPFGKNGGVDN"  
complement(8440..11121)  
/note="potential pseudogene, contains 3 orfs which are  
separated by single stop codons and which share high

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

gene

CDS

LTR

gene

CDS

misc\_feature

homology to SPAC8A4.02c but with more (about 14) 36 aa repeats"

gene complement(9025..9558)  
/genes="SPAC8A4.04c"  
CDS complement(9025..9558)  
/genes="SPAC8A4.04c"  
/note="Protein sequence is in conflict with the conceptual translation; SPAC8A4.04c, unknown orf, len: 177, possibly spliced but the three separate reading frames in this region are only separated by single stop codons and all share high similarity to SPAC8A4.02c. Therefore these 3 orfs could be a pseudogene"  
/codon\_start=1  
/product="unknown"  
/protein\_id="CAA91514.1"  
/db\_xref="GI:1376031"  
/translation="FADSVSRVEVIEPIAGTLTSTVYSGSGFTTTLTGASGVSGTVEVQPTAGTITTVYSGTQGTTLTGTASGSVGTVEVQPTAGTITTVTFEGSSTFTVLNLSAGSVSGTVDVVKAGPNTIYSGTPTTVYSSASSAATVVVPIPTAVCSGIRGLQYAVNYNTIASSKS"  
gene complement(9559..11121)  
/genes="SPAC8A4.05c"  
CDS complement(9559..11121)  
/genes="SPAC8A4.05c"  
/note="SPAC8A4.05c, unknown orf, len: 520, possibly spliced to downstream orf but the three separate reading frames in this region are only separated by single stop codons and all share high similarity to SPAC8A4.02c. Therefore these 3 orfs could be a pseudogene"  
/codon\_start=1  
/product="unknown"  
/protein\_id="CAA91515.1"  
/db\_xref="GI:1052536"  
/db\_xref="SWISS-PROT:Q09881"  
/translation="MFSEFRSTLVFLSIFTVLSEPRVSIGLAKRTISQTSSSSLYTCPDYITFSEGSEYTTIYPSKSSSVANHSITRTIDSGTIATYFTFLPDGEI VIRDIEPVAKTLLTITSGSLELTTLTATASGTVEVVEPLAGTATVTVYSGSVE HNTTLATASGTVEVVEPTAGTITTVYSGDQETTLATASGAVPTVEVDPA AGTVTTIYSGSVYNTLTATASETLPGTIEVVEPLAGTITTVYSGSAQYETTLATA SCTVSGTVEVPEPAAGTLTTVYSGSEATTVASGTVEVETAGTVYQTSCMDLPL TSLTVATASGTPVGLVLTAPFAGTITTVYSGSQYTTTVATAAGCVSGVEVVEP TGTVTVTSYSGSAQYETTLAEASGTVPGTVEVETAGTITTVYSGDQETTLAT ASGAVPTVEVVDPAAGTITTVYSGQLFTTALATATGSIPTVEIVDPAAGTTSIT TVSGSMETSTL"  
gene join(13149..13177,13221..14772,14821..14913)  
/genes="SPAC8A4.06"  
CDS join(13149..13177,13221..14772,14821..14913)  
/note="SPAC8A4.06, similar to BX42\_DROME P39736 puff specific protein BX42, (46.0% identity in 409 aa overlap)"  
/codon\_start=1  
/product="unknown"

alignment\_scores:  
Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-701-711-2 x SPAC8A4

Align seg 1/1 to: SPAC8A4 from: 1 to: 38141

237 LeuAsnAlaGlyPheValArgPheGlu 245

|||||

11145 TTAACGGCGGCTTGAAGATTGAA 11171

seq\_name: gb\_pl:SPCC584

seq\_documentation\_block:

LOCUS SPCC584

DEFINITION S.pombe chromosome III cosmid c584.

ACCESSION AL032824

VERSION  
KEYWORDS

AL032824.2 GI:4456829  
amino-acid permease; Chromatin binding; FAD/NAD-binding domain;  
LTR; metal-binding regulatory protein; replication factor-a protein  
2; ribonuclease; sec1 family; secretory protein; septin homolog;  
snw family nuclear protein; sulfite reductase; syntaxin binding  
protein.

SOURCE  
ORGANISM

Schizosaccharomyces pombe  
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
Schizosaccharomycetales; Schizosaccharomycetaceae;  
Schizosaccharomycetes.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

1 (bases 1 to 41803)  
Direct Submission  
Submitted (29-OCT-1998) European Schizosaccharomycetes genome  
sequencing project, Sanger Centre, The Wellcome Trust Genome  
Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrellesanger.ac.uk  
On Mar 21, 1999 this sequence version replaced gi:3819700.

COMMENT

Notes:  
Details of yeast sequencing at the Sanger Centre are available on  
the World Wide Web.  
URL, [http://www.sanger.ac.uk/Projects/S\\_pombe/](http://www.sanger.ac.uk/Projects/S_pombe/)  
During 1995 to 1996 about 66% of S. pombe chromosome 1 was  
sequenced by the Sanger Centre. The sequencing of the S. pombe  
genome is now being continued with funding from The European  
Commission. Fourteen European sequencing laboratories, including  
the Sanger Centre, are participating in the project.  
Protein coding regions (CDS) have been predicted with the help of  
computer analysis using the genefinder program in Pombase (an ACEDB  
database) with additional predictions for the branch-acceptor sites  
supplied by the program Sp3splice. CAUTION: It is possible that for  
any individual CDS we may have underestimated or overestimated the  
number of introns/exons or we may not have chosen the correct  
splice donor/acceptor sites. CDS are numbered using the following  
system eg SPCC35H2.01c. SP (S. pombe), B (chromosome 2), c25H2  
(cosmid name), .01 (first CDS), c (complementary strand).  
The more significant matches with motifs in the PROSITE database  
are also included but some of these may be fortuitous. The length  
in codons is given for each CDS.  
IMPORTANT: This sequence MAY NOT be the entire insert of the  
sequenced clone. It may be shorter because we only sequence  
overlapping sections once, or longer, because we arrange for a  
small overlap between neighbouring submissions. c584 overlaps c188  
at its 5' end, and c1753 at its 3' end.

FEATURES  
source

1..41803  
/organism="Schizosaccharomycetes pombe"  
/strain="972h-"  
/db\_xref="taxon:4896"  
/chromosome="III"  
/map="IIIR"

gene

/clone="cosmid c584"  
complement(1..649)  
/gene="SPCC584.07c"  
complement(1..649)  
/partial

CDS

/gene="SPCC584.07c"  
/note="SPCC584.07c, len:<216: contains 36 aa repeat"  
/codon\_start=1  
/label="SPCC584.07c  
/product="hypothetical repeat containing protein"  
/protein\_id="CAB37420.1"  
/db\_xref="GI:4456830"

/translation="MFSEFRSTLVFLSIFTVLSEPRVSIGLAKRTISQTSSSSLY  
TCPDYITFSEGSEYTTIYPSKSSSVANHSITRTIDSGTIATYFTFLPDGEI  
VIRDIEPVAKTLLTITSGSLELTTLTATASGTVEVVEPLAGTATVTVYSGSVE  
HNTTLATASGTVEVVEPTAGTITTVYSGDQETTLATASGAVPTVEVDPA  
AGTVTTIYSGSVYNTLTATASETLPGTIEVVEPLAGTITTVYSGSAQYETTLATA  
SCTVSGTVEVPEPAAGTLTTVYSGSEATTVASGTVEVETAGTVYQTSCMDLPL  
TSLTVATASGTPVGLVLTAPFAGTITTVYSGSQYTTTVATAAGCVSGVEVVEP  
TGTVTVTSYSGSAQYETTLAEASGTVPGTVEVETAGTITTVYSGDQETTLAT  
ASGAVPTVEVVDPAAGTITTVYSGQLFTTALATATGSIPTVEIVDPAAGTTSIT  
TVSGSMETSTL"  
1..8113

misc\_feature

/note="overlap with c188 S. pombe chromosome 3"

misc\_feature

2047..2159  
/note="small conserved intergenic region, possible rRNA,  
2 3532 23644 c188 S. pombe chromosome 3"  
join(2677..2705,2749..4300,4349..4441)

gene





## KEYWORDS

HTG; HTGS\_PHASE0.

## SOURCE

house mouse.

## ORGANISM

Mus musculus

## REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## AUTHORS

1 (bases 1 to 64341)

## TITLE

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

## JOURNAL

Mus musculus, clone RP24-319020

## REFERENCE

Unpublished

## AUTHORS

2 (bases 1 to 64341)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Bouckgalter,B., Brown,A., Camarata,J., Campoliano,A., Chang,J., Chazaro,B., Choepel,Y., Collangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Glade,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kellis,C., LaRocque,K., Lamazeas,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

## TITLE

Direct Submission

## JOURNAL

Submitted (23-NOV-2001)

## COMMENT

Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WITR

Web site: <http://www-seq.wi.mit.edu>Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L20053

Center clone name: 319\_Q\_20

-----

\* NOTE: This record contains 82 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 676: contig of 676 bp in length  
\* 677 776: gap of 100 bp  
\* 777 1447: contig of 671 bp in length  
\* 1448 1547: gap of 100 bp  
\* 1548 2204: contig of 657 bp in length  
\* 2205 2304: gap of 100 bp  
\* 2305 2971: contig of 667 bp in length  
\* 2972 3071: gap of 100 bp  
\* 3072 3689: contig of 618 bp in length  
\* 3690 3789: gap of 100 bp  
\* 3790 4489: contig of 700 bp in length  
\* 4490 4589: gap of 100 bp  
\* 4590 5287: contig of 698 bp in length  
\* 5288 5387: gap of 100 bp  
\* 5388 6073: contig of 686 bp in length

6074 6173: gap of 100 bp  
\* 6174 6843: contig of 670 bp in length  
\* 6844 6943: gap of 100 bp  
\* 6944 7584: contig of 641 bp in length  
\* 7585 7684: gap of 100 bp  
\* 7685 8386: contig of 702 bp in length  
\* 8387 8486: gap of 100 bp  
\* 8487 9164: contig of 678 bp in length  
\* 9165 9264: gap of 100 bp  
\* 9265 9947: contig of 683 bp in length  
\* 9948 10047: gap of 100 bp  
\* 10048 10755: contig of 708 bp in length  
\* 10756 10855: gap of 100 bp  
\* 10856 11554: contig of 699 bp in length  
\* 11555 11654: gap of 100 bp  
\* 11655 12351: contig of 697 bp in length  
\* 12352 12451: gap of 100 bp  
\* 12452 13145: contig of 694 bp in length  
\* 13146 13245: gap of 100 bp  
\* 13246 13906: contig of 661 bp in length  
\* 13907 14006: gap of 100 bp  
\* 14007 14690: contig of 684 bp in length  
\* 14691 14790: gap of 100 bp  
\* 14791 15484: contig of 694 bp in length  
\* 15485 15584: gap of 100 bp  
\* 15585 16279: contig of 695 bp in length  
\* 16280 16379: gap of 100 bp  
\* 16380 17068: contig of 689 bp in length  
\* 17069 17168: gap of 100 bp  
\* 17169 17855: contig of 687 bp in length  
\* 17856 17955: gap of 100 bp  
\* 17956 18636: contig of 681 bp in length  
\* 18637 18736: gap of 100 bp  
\* 18737 19424: contig of 688 bp in length  
\* 19425 19524: gap of 100 bp  
\* 19525 20243: contig of 719 bp in length  
\* 20244 20343: gap of 100 bp  
\* 20344 21059: contig of 716 bp in length  
\* 21060 21159: gap of 100 bp  
\* 21160 21865: contig of 706 bp in length  
\* 21866 21965: gap of 100 bp  
\* 21966 22664: contig of 699 bp in length  
\* 22665 22764: gap of 100 bp  
\* 22765 23440: contig of 676 bp in length  
\* 23441 23540: gap of 100 bp  
\* 23541 24221: contig of 681 bp in length  
\* 24222 24321: gap of 100 bp  
\* 24322 25010: contig of 689 bp in length  
\* 25011 25110: gap of 100 bp  
\* 25111 25812: contig of 702 bp in length  
\* 25813 25912: gap of 100 bp  
\* 25913 26615: contig of 703 bp in length  
\* 26616 26715: gap of 100 bp  
\* 26716 27390: contig of 675 bp in length  
\* 27391 27490: gap of 100 bp  
\* 27491 28202: contig of 712 bp in length  
\* 28203 28302: gap of 100 bp  
\* 28303 28994: contig of 692 bp in length  
\* 28995 29094: gap of 100 bp  
\* 29095 29772: contig of 678 bp in length  
\* 29773 29872: gap of 100 bp  
\* 29873 30580: contig of 708 bp in length  
\* 30581 30680: gap of 100 bp  
\* 30681 31369: contig of 689 bp in length  
\* 31370 31469: gap of 100 bp  
\* 31470 32134: contig of 665 bp in length  
\* 32135 32234: gap of 100 bp  
\* 32235 32923: contig of 689 bp in length  
\* 32924 33023: gap of 100 bp  
\* 33024 33712: contig of 689 bp in length  
\* 33713 33812: gap of 100 bp  
\* 33813 34511: contig of 699 bp in length  
\* 34512 34611: gap of 100 bp

```

* 34612 35283: contig of 672 bp in length
* 35284 35383: gap of 100 bp
* 35384 36105: contig of 722 bp in length
* 36106 36205: gap of 100 bp
* 36206 36884: contig of 679 bp in length
* 36885 36984: gap of 100 bp
* 36985 37658: contig of 674 bp in length
* 37659 37758: gap of 100 bp
* 37759 38414: contig of 656 bp in length
* 38415 38514: gap of 100 bp
* 38515 39196: contig of 682 bp in length
* 39197 39296: gap of 100 bp
* 39297 40003: contig of 707 bp in length
* 40004 40103: gap of 100 bp
* 40104 40772: contig of 669 bp in length
* 40773 40872: gap of 100 bp
* 40873 41560: contig of 688 bp in length
* 41561 41660: gap of 100 bp
* 41661 42309: contig of 649 bp in length
* 42310 42409: gap of 100 bp
* 42410 43115: contig of 706 bp in length
* 43116 43215: gap of 100 bp
* 43216 43908: contig of 693 bp in length
* 43909 44008: gap of 100 bp
* 44009 44702: contig of 694 bp in length
* 44703 44802: gap of 100 bp
* 44803 45498: contig of 696 bp in length
* 45499 45598: gap of 100 bp
* 45599 46281: contig of 683 bp in length
* 46282 46381: gap of 100 bp
* 46382 47167: contig of 686 bp in length
* 47168 47830: contig of 663 bp in length
* 47831 47930: gap of 100 bp
* 47931 48611: contig of 681 bp in length
* 48612 48711: gap of 100 bp
* 48712 49396: contig of 685 bp in length
* 49397 49496: gap of 100 bp
* 49497 50186: contig of 690 bp in length
* 50187 50286: gap of 100 bp
* 50287 50956: contig of 670 bp in length
* 50957 51056: gap of 100 bp
* 51057 51767: contig of 711 bp in length
* 51768 51867: gap of 100 bp
* 51868 52565: contig of 698 bp in length
* 52566 52665: gap of 100 bp
* 52666 53364: contig of 699 bp in length
* 53365 53464: gap of 100 bp
* 53465 54141: contig of 677 bp in length
* 54142 54241: gap of 100 bp

```

## alignment\_scores:

```

Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

## alignment\_block:

```

US-09-701-711-2 x AC102899/rev ..

```

```

Align seg 1/1 to reverse of: AC102899 from: 1 to: 64341

```

```

607 LeuThrValGlyPheGlyAspLysThr 615
|||||

```

```

48914 CTCACGGTTGGTTGGGACAGACG 48888

```

```

seq_name: gb_htg:AC095320

```

```

seq_documentation_block:

```

```

LOCUS AC095320 94948 bp DNA linear HTG 20-DEC-2001
DEFINITION Rattus norvegicus clone CH230-141D12, *** SEQUENCING IN PROGRESS
*** 57 unordered pieces.
ACCESSION AC095320
VERSION AC095320.2 GI:17956648

```

## KEYWORDS

```

SOURCE HTGS_PHASE1.

```

## ORGANISM

```

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

```

## REFERENCE

```

1 (bases 1 to 94948)

```

## AUTHORS

```

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralungu,H.C., Are,J.R., Banks,T., Barbaria,J.,
Benton,J., Binage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homsif,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Louiseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
Ogih,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,F., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

```

## TITLE

```

JOURNAL Direct Submission

```

## REFERENCE

```

2 (bases 1 to 94948)

```

## AUTHORS

```

Worley,K.C.

```

## JOURNAL

```

Submitted (16-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:15625874.

```

## COMMENT

```

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GFXY
Center clone name: CH230-141D12
----- Summary Statistics
Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 67588 bases at least Q40
Consensus quality: 73195 bases at least Q30
Consensus quality: 77056 bases at least Q20
Estimated insert size: 31978; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-tp estimation
Quality coverage: 0.4x in Q20 bases; sum-of-contigs estimation
-----

```

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 57 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 3273: contig of 3273 bp in length  
\* 3274: gap of unknown length  
\* 3374: contig of 2037 bp in length  
\* 5411: gap of unknown length  
\* 5511: contig of 2134 bp in length  
\* 7645: gap of unknown length  
\* 7744: contig of 1012 bp in length  
\* 8756: gap of unknown length  
\* 8857: contig of 2117 bp in length  
\* 10973: contig of 1173 bp in length  
\* 10974: contig of 2257 bp in length  
\* 13330: contig of 1485 bp in length  
\* 13331: contig of 1485 bp in length  
\* 14915: contig of 1485 bp in length  
\* 15015: gap of unknown length  
\* 15016: contig of 2683 bp in length  
\* 17698: gap of unknown length  
\* 17699: contig of 1741 bp in length  
\* 19539: gap of unknown length  
\* 19540: contig of 2878 bp in length  
\* 19640: contig of 2878 bp in length  
\* 22518: gap of unknown length  
\* 22617: contig of 1821 bp in length  
\* 24438: gap of unknown length  
\* 24439: contig of 1933 bp in length  
\* 24539: gap of unknown length  
\* 26471: contig of 1132 bp in length  
\* 26571: gap of unknown length  
\* 26572: contig of 1132 bp in length  
\* 27703: gap of unknown length  
\* 27803: contig of 2090 bp in length  
\* 29893: gap of unknown length  
\* 29894: contig of 2268 bp in length  
\* 32261: gap of unknown length  
\* 32262: contig of 1456 bp in length  
\* 32362: gap of unknown length  
\* 33818: contig of 2306 bp in length  
\* 33918: gap of unknown length  
\* 36223: contig of 1422 bp in length  
\* 36323: gap of unknown length  
\* 36324: contig of 1422 bp in length  
\* 37446: gap of unknown length  
\* 37846: contig of 1372 bp in length  
\* 39217: gap of unknown length  
\* 39318: contig of 1464 bp in length  
\* 40782: gap of unknown length  
\* 40882: contig of 1985 bp in length  
\* 42867: gap of unknown length  
\* 42967: contig of 1441 bp in length  
\* 44408: gap of unknown length  
\* 44508: contig of 1090 bp in length  
\* 45598: gap of unknown length  
\* 45698: contig of 1235 bp in length  
\* 46933: gap of unknown length  
\* 47033: contig of 1008 bp in length  
\* 48040: gap of unknown length  
\* 48140: contig of 1568 bp in length  
\* 49708: gap of unknown length  
\* 49709: contig of 1629 bp in length  
\* 49809: gap of unknown length  
\* 51438: contig of 1804 bp in length  
\* 51538: gap of unknown length  
\* 53342: contig of 1294 bp in length  
\* 53442: gap of unknown length  
\* 54736: contig of 1435 bp in length  
\* 54836: gap of unknown length  
\* 56271: contig of 1048 bp in length  
\* 56371: gap of unknown length  
\* 57519: contig of 1102 bp in length  
\* 58620: contig of 1102 bp in length

\* 58621: gap of unknown length  
\* 58721: contig of 1356 bp in length  
\* 60077: gap of unknown length  
\* 60176: contig of 2197 bp in length  
\* 62374: gap of unknown length  
\* 62474: contig of 1452 bp in length  
\* 63926: gap of unknown length  
\* 64026: contig of 1601 bp in length  
\* 65627: gap of unknown length  
\* 65727: contig of 1917 bp in length  
\* 67644: gap of unknown length  
\* 67744: contig of 1406 bp in length  
\* 69150: gap of unknown length  
\* 69249: contig of 1094 bp in length  
\* 70344: gap of unknown length  
\* 70444: contig of 1876 bp in length  
\* 72120: gap of unknown length  
\* 72220: contig of 1199 bp in length  
\* 73419: gap of unknown length  
\* 73519: contig of 1236 bp in length  
\* 74755: gap of unknown length  
\* 74855: contig of 1259 bp in length  
\* 76114: gap of unknown length  
\* 76214: contig of 1847 bp in length  
\* 78061: gap of unknown length  
\* 78161: contig of 1413 bp in length  
\* 79574: gap of unknown length  
\* 79674: contig of 1117 bp in length  
\* 80791: gap of unknown length  
\* 80891: contig of 1288 bp in length  
\* 82179: gap of unknown length  
\* 82279: contig of 1009 bp in length  
\* 83288: gap of unknown length  
\* 83388: contig of 1284 bp in length  
\* 84672: gap of unknown length  
\* 84772: contig of 1431 bp in length  
\* 86203: gap of unknown length  
\* 86303: contig of 1183 bp in length  
\* 87486: gap of unknown length  
\* 87586: contig of 1123 bp in length  
\* 88709: gap of unknown length  
\* 88809: contig of 1351 bp in length  
\* 90160: gap of unknown length  
\* 90260: contig of 1225 bp in length  
\* 91485: gap of unknown length  
\* 91585: contig of 1106 bp in length  
\* 92691: gap of unknown length  
\* 92791: contig of 1025 bp in length  
\* 93816: gap of unknown length

## alignment\_scores:

Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-701-711-2 x AC095320/rev ..

Align seg 1/1 to reverse of: AC095320 from: 1 to: 94948

679 AlaArgGlyGlnGlnThrLeu 687  
|||||  
29093 GCCAGAGAGGGCAGCAGACCATTTG 29067

seq\_name: gb\_pr:AL359454

seq\_documentation\_block:

LOCUS AL359454 101830 bp DNA linear PRI 30-MAY-2001  
DEFINITION Human DNA sequence from clone RP11-97E23 on chromosome 13, complete  
sequence.  
ACCESSION AL359454  
VERSION AL359454.7 GI:14272304  
KEYWORDS HTG.

SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT	repeat_region
human.	Homo sapiens	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 101830)	Smith, M.	Direct Submission	Submitted (30-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk	requests: clonerequest@sanger.ac.uk	repeat_region
					On May 31, 2001 this sequence version replaced gi:13398783.	During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.	repeat_region
					This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at	repeat_region	repeat_region
					http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at	repeat_region	repeat_region
					http://www.sanger.ac.uk/HGP/Chr13	repeat_region	repeat_region
					RP11-97E23 is from the library RPCR-11.1 constructed by the group of Pieter de Jong. For further details see	repeat_region	repeat_region
					http://www.chori.org/bacpac/home.htm	repeat_region	repeat_region
					VECTOR: pBACE3.6	repeat_region	repeat_region
					IMPORTANT: This sequence is not the entire insert of clone RP11-97E23 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap	repeat_region	repeat_region
					The true right end of clone RP11-97E23 is at 101830 in this sequence. The true left end of clone RP11-161P17 is at 78762 in this sequence. The true right end of clone RP11-57H24 is at 100 in this sequence.	repeat_region	repeat_region
FEATURES	Location/Qualifiers						repeat_region
source	1..101830						repeat_region
	/organism="Homo sapiens"						repeat_region
	/db_xref="taxon:9606"						repeat_region
	/chromosome="13"						repeat_region
	/clone="RP11-97E23"						repeat_region
	/clone_lib="RPCR-11.1"						repeat_region
	671..980						repeat_region
repeat_region	/note="AluX repeat: matches 1..310 of consensus"						repeat_region
repeat_region	2541..2808						repeat_region
repeat_region	/note="AluJo repeat: matches 34..296 of consensus"						repeat_region
repeat_region	3048..3084						repeat_region
repeat_region	/note="MIR repeat: matches 156..192 of consensus"						repeat_region
repeat_region	3577..3713						repeat_region
repeat_region	/note="MIR repeat: matches 4..143 of consensus"						repeat_region
repeat_region	4667..4753						repeat_region
repeat_region	/note="HY3 repeat: matches 1..99 of consensus"						repeat_region
repeat_region	4756..5060						repeat_region
repeat_region	/note="AluSg repeat: matches 1..305 of consensus"						repeat_region
repeat_region	5291..5428						repeat_region
repeat_region	/note="LTR16C repeat: matches 248..381 of consensus"						repeat_region
repeat_region	5429..5740						repeat_region
repeat_region	/note="AluX repeat: matches 1..312 of consensus"						repeat_region
repeat_region	5741..5896						repeat_region
repeat_region	/note="LTR16C repeat: matches 79..248 of consensus"						repeat_region
repeat_region	6921..7101						repeat_region
repeat_region	/note="MIR repeat: matches 23..215 of consensus"						repeat_region
repeat_region	7357..7669						repeat_region
repeat_region	/note="AluSg repeat: matches 1..312 of consensus"						repeat_region
repeat_region	8167..8297						repeat_region
repeat_region	/note="FLAM_A repeat: matches 4..133 of consensus"						repeat_region
repeat_region	8385..8724						repeat_region
repeat_region	/note="170 copies 2 mer at 73% conserved"						repeat_region
repeat_region	8725..8924						repeat_region
repeat_region	/note="100 copies 2 mer ta 80% conserved"						repeat_region
repeat_region	10158..10293						repeat_region
repeat_region	/note="L1MC2 repeat: matches 6079..6217 of consensus"						repeat_region
repeat_region	10294..10471						repeat_region
repeat_region	/note="AluJo repeat: matches 121..298 of consensus"						repeat_region
repeat_region	10505..10589						repeat_region
repeat_region	/note="L1MC2 repeat: matches 6246..6328 of consensus"						repeat_region
repeat_region	12182..12493						repeat_region
repeat_region	/note="AluJo repeat: matches 1..302 of consensus"						repeat_region
repeat_region	13412..13535						repeat_region
repeat_region	/note="FLAM_C repeat: matches 3..126 of consensus"						repeat_region
repeat_region	14042..14232						repeat_region
repeat_region	/note="MIR repeat: matches 2..219 of consensus"						repeat_region
repeat_region	14950..15247						repeat_region
repeat_region	/note="AluX repeat: matches 1..300 of consensus"						repeat_region
repeat_region	16669..16962						repeat_region
repeat_region	/note="AluJo repeat: matches 1..298 of consensus"						repeat_region
repeat_region	17910..18230						repeat_region

```
/note="L2 repeat: matches 2069. .2419 of consensus"
32005. .32317
/note="AlusX repeat: matches 1. .307 of consensus"
32812. .32940
/note="FLAM_A repeat: matches 16. .140 of consensus"
33242. .33404
/note="MIR repeat: matches 84. .261 of consensus"
33747. .34249
/note="CpG island"
/evidence-not_experimental
35100. .35458
/note="MER47A repeat: matches 3. .366 of consensus"
35487. .35554
/note="MIR repeat: matches 67. .137 of consensus"
35754. .35850
/note="L2 repeat: matches 2005. .2109 of consensus"
36054. .36356
/note="AluJb repeat: matches 1. .300 of consensus"
36447. .36511
/note="L2 repeat: matches 2679. .2743 of consensus"
37117. .37963
/note="MIR repeat: matches 2. .262 of consensus"
38176. .38482
/note="AlusP repeat: matches 1. .308 of consensus"
39199. .39308
/note="L1MD2 repeat: matches 6014. .6122 of consensus"
41473. .41584
/note="MIR repeat: matches 23. .155 of consensus"
41678. .41915
/note="MER20 repeat: matches 1. .214 of consensus"
42119. .42278
/note="FAM repeat: matches 1. .168 of consensus"
43498. .43669
/note="L2 repeat: matches 742. .919 of consensus"
43708. .43769
/note="L2 repeat: matches 1724. .1783 of consensus"
43807. .44242
/note="L2 repeat: matches 958. .1327 of consensus"
44243. .44504
/note="AluJo repeat: matches 49. .306 of consensus"
44505. .44698
/note="L2 repeat: matches 1327. .1490 of consensus"
44699. .44993
/note="AluJb repeat: matches 3. .293 of consensus"
44994. .45418
/note="L2 repeat: matches 1490. .2173 of consensus"
45638. .45929
/note="Alusq repeat: matches 1. .292 of consensus"

alignment_scores:
  Quality: 9.00 Length: 9
  Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AL359454/rev ..
Align seg 1/1 to reverse of: AL359454 from: 1 to: 101830

502 TyrAspAsnLysAsnIleSerAsnTyr 510
|||||
57836 TATGATAACAAAAATATCTCAAAATTAT 57810

seq_name: gb_htg:AC094530
seq_documentation_block:
LOCUS AC094530
DEFINITION Rattus norvegicus clone CH230-4L15, *** SEQUENCING IN PROGRESS ***,
53 unordered pieces.
ACCESSION AC094530
VERSION AC094530.3 GI:17941265
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
```

```
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 106997)
REFERENCE
AUTHORS
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Albrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbacia,J.,
Benton,J., Bimege,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,
Hollins,B., Homsif,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulsegh,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
Ogih,M., Okunou,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Perry,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N.,
Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Thomas,N.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 106997)
Worley,K.C.
Direct Submission
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:15799323.
-----
Center: Baylor College of Medicine
Genome Center
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-----
Project Information
Center project name: GAVE
Center clone name: CH230-4L15
-----
Summary Statistics
Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 78042 bases at least Q40
Consensus quality: 86842 bases at least Q30
Consensus quality: 93110 bases at least Q20
Estimated insert size: 79937; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 0.9x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
```

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 53 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 4485: contig of 4485 bp in length  
\* 4486 4585: gap of unknown length  
\* 4586 10783: contig of 6198 bp in length  
\* 10784 10883: gap of unknown length  
\* 10884 13167: contig of 2284 bp in length  
\* 13168 13267: gap of unknown length  
\* 13268 16560: contig of 3293 bp in length  
\* 16561 16660: gap of unknown length  
\* 16661 19562: contig of 2902 bp in length  
\* 19563 19662: gap of unknown length  
\* 19663 21931: contig of 2289 bp in length  
\* 21932 22031: gap of unknown length  
\* 22032 25291: contig of 3260 bp in length  
\* 25292 25391: gap of unknown length  
\* 25392 26888: contig of 1497 bp in length  
\* 26889 26988: gap of unknown length  
\* 26989 29002: contig of 2014 bp in length  
\* 29003 29102: gap of unknown length  
\* 29103 31586: contig of 2484 bp in length  
\* 31587 31686: gap of unknown length  
\* 31687 34451: contig of 2765 bp in length  
\* 34452 34551: gap of unknown length  
\* 34552 37275: contig of 2724 bp in length  
\* 37276 37375: gap of unknown length  
\* 37376 39363: contig of 1988 bp in length  
\* 39364 39463: gap of unknown length  
\* 39464 41942: contig of 2479 bp in length  
\* 41943 42042: gap of unknown length  
\* 42043 44545: contig of 2503 bp in length  
\* 44546 44645: gap of unknown length  
\* 44646 46114: contig of 1769 bp in length  
\* 46115 46314: gap of unknown length  
\* 46315 48423: contig of 1909 bp in length  
\* 48424 48523: gap of unknown length  
\* 48524 50116: contig of 1593 bp in length  
\* 50117 50216: gap of unknown length  
\* 50217 52249: contig of 2033 bp in length  
\* 52250 52349: gap of unknown length  
\* 52350 54385: contig of 2036 bp in length  
\* 54386 54485: gap of unknown length  
\* 54486 56136: contig of 1651 bp in length  
\* 56137 56236: gap of unknown length  
\* 56237 57268: contig of 1032 bp in length  
\* 57269 57368: gap of unknown length  
\* 57369 59226: contig of 1858 bp in length  
\* 59227 59326: gap of unknown length  
\* 59327 60774: contig of 1448 bp in length  
\* 60775 60874: gap of unknown length  
\* 60875 63429: contig of 2555 bp in length  
\* 63430 63529: gap of unknown length  
\* 63530 65824: contig of 2295 bp in length  
\* 65825 65924: gap of unknown length  
\* 65925 67471: contig of 1547 bp in length  
\* 67472 67571: gap of unknown length  
\* 67572 69443: contig of 1872 bp in length  
\* 69444 69543: gap of unknown length  
\* 69544 71798: contig of 2255 bp in length  
\* 71799 71898: gap of unknown length  
\* 71899 73517: contig of 1619 bp in length  
\* 73518 73617: gap of unknown length  
\* 73618 75617: contig of 2000 bp in length  
\* 75618 75717: gap of unknown length  
\* 75718 77253: contig of 1536 bp in length  
\* 77254 77353: gap of unknown length  
\* 77354 79220: contig of 1867 bp in length

\* 79221 79320: gap of unknown length  
\* 79321 80763: contig of 1443 bp in length  
\* 80764 82206: gap of unknown length  
\* 82207 82306: contig of 1343 bp in length  
\* 82307 84307: gap of unknown length  
\* 84308 84407: gap of unknown length  
\* 84408 85700: contig of 1463 bp in length  
\* 85701 85971: gap of unknown length  
\* 85972 87046: contig of 1076 bp in length  
\* 87047 87146: gap of unknown length  
\* 87147 88524: contig of 1378 bp in length  
\* 88525 88624: gap of unknown length  
\* 88626 89791: contig of 1167 bp in length  
\* 89792 89891: gap of unknown length  
\* 89892 91038: contig of 1147 bp in length  
\* 91039 91138: gap of unknown length  
\* 91139 92175: contig of 1037 bp in length  
\* 92176 92275: gap of unknown length  
\* 92276 93492: contig of 1217 bp in length  
\* 93493 93592: gap of unknown length  
\* 93593 94771: contig of 1179 bp in length  
\* 94772 94871: gap of unknown length  
\* 94872 95994: contig of 1123 bp in length  
\* 95995 96094: gap of unknown length  
\* 96095 97483: contig of 1389 bp in length  
\* 97484 97583: gap of unknown length  
\* 97584 98909: contig of 1226 bp in length  
\* 98910 98909: gap of unknown length  
\* 98911 100242: contig of 1333 bp in length  
\* 100243 100342: gap of unknown length  
\* 100343 101569: contig of 1227 bp in length  
\* 101570 101669: gap of unknown length  
\* 101670 103028: contig of 1359 bp in length  
\* 103029 103128: gap of unknown length  
\* 103129 104159: contig of 1031 bp in length  
\* 104160 104259: gap of unknown length  
\* 104260 105435: contig of 1176 bp in length  
\* 105436 105535: gap of unknown length  
\* 105536 106997: contig of 1462 bp in length.

## FEATURES

source  
1..106997  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/clone="CH230-4L15"

BASE COUNT 30338 a 20613 c 20041 g 30563 t 5442 others  
ORIGIN

## alignment\_scores:

Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-701-711-2 x AC094530 ..

Align seg 1/1 to: AC094530 from: 1 to: 106997

632 LysSerValLeuArgGlyTyrAlaLys 640  
|||||  
101456 AAATCTGTACTAAGAGGATATGCAAAA 101482

seq\_name: gb\_htg:AC087596

## seq\_documentation\_block:

LOCUS AC087596 109210 bp DNA linear HTG 12-JAN-2001  
DEFINITION Oryza sativa chromosome 1 clone OSJNBa0055C02, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 3 ordered pieces.  
ACCESSION AC087596  
VERSION AC087596.1 GI:12084798  
KEYWORDS HTG; HTGS-PHASE2.  
SOURCE Oryza sativa.  
ORGANISM Oryza sativa

Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzeae; Oryza.

REFERENCE 1 (bases 1 to 109210)

AUTHORS Lee, J.-S., Hahn, J.-H., Lee, M.-C., Yoon, U.-H., Yun, D.-W., Kim, H.-I.  
and Eun, M.-Y.

TITLE Oryza sativa BAC OSJNBa0055C02 genomic sequence

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 109210)

AUTHORS Hahn, J.-H. and Eun, M.-Y.

TITLE Direct Submission

JOURNAL Submitted (12-JAN-2001) Rice Genome Sequencing Project, National

Institute of Agricultural Science and Technology (NIAT), RDA, 249

Seodun-dong, Suwon 441-707, Korea (E-mail: jhahn@rda.go.kr,

Tel: 82-31-290-0309, Fax: 82-31-290-0308)

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 3 contigs. Gaps between the contigs

\* are represented as runs of N. The order of the pieces

\* is believed to be correct as given, however the sizes

\* of the gaps between them are based on estimates that have

\* provided by the submitter.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.

\* 1 25874: contig of 25874 bp in length

\* 25875 25974: gap of unknown length

\* 25975 105215: contig of 79241 bp in length

\* 105216 105315: gap of unknown length

\* 105316 109210: contig of 3895 bp in length.

FEATURES Location/Qualifiers

source 1..109210

/organism="Oryza sativa"

/cultivar="Nipponbare"

/db\_xref="taxon:4530"

/chromosome="1"

/clone="OSJNBa0055C02"

BASE COUNT 31001 a 23200 c 23520 g 31287 t 202 others

ORIGIN

alignment\_scores:

Quality: 9.00 Length: 9

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-701-711-2 x AC087596 ..

Align seg 1/1 to: AC087596 from: 1 to: 109210

514 SerTyrGlyGlySerLeuSerTyrGly 522

|||||

15087 TCCTACGGGGATCATTCCTACGGT 15113

seq\_name: gb\_htg:LMFLCHR16\_09

seq\_documentation\_block:

WPCOMMENT

Sequence split into 11 fragments LOCUS LMFLCHR16 Accession AL499619

Fragment Name Begin End

LMFLCHR16\_00 1 110000

LMFLCHR16\_01 100001 210000

LMFLCHR16\_02 200001 310000

LMFLCHR16\_03 300001 410000

LMFLCHR16\_04 400001 510000

LMFLCHR16\_05 500001 610000

LMFLCHR16\_06 600001 710000

LMFLCHR16\_07 700001 810000

LMFLCHR16\_08 800001 910000

LMFLCHR16\_09 900001 1010000

LMFLCHR16\_10 1000001 1030105

Continuation (10 of 11) of LMFLCHR16 from base 900001 (AL499619 Leishmania major chromos

alignment\_scores:

Quality: 9.00 Length: 9

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-701-711-2 x LMFLCHR16\_09/rev ..

Align seg 1/1 to reverse of: LMFLCHR16\_09 from: 1 to: 110000

287 GlnAlaGluLeuGluAlaLeuLeuLys 295

|||||

76655 CAGGCCGAGCTTGAGCGCTGCTGAAG 76629

seq\_name: gb\_htg:AC016033

seq\_documentation\_block:

LOCUS AC016033

DEFINITION Homo sapiens clone RP11-13024, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 11

unordered pieces.

ACCESSION AC016033

VERSION AC016033.5 GI:17977644

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_FULLTOP; HTGS\_ACTIVEFIN.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 117693)

AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.

TITLE Homo sapiens chromosome, clone RP11-13024

REFERENCE 2 (bases 1 to 117693)

AUTHORS Unpublished

JOURNAL

REFERENCE

AUTHORS

TITLE

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

```

* 19507 35142: contig of 15636 bp in length
* 35143 35242: gap of 100 bp
* 35243 40193: contig of 4951 bp in length
* 40194 40293: gap of 100 bp
* 40294 52911: contig of 12618 bp in length
* 52912 53011: gap of 100 bp
* 53012 57184: contig of 4173 bp in length
* 57185 57284: gap of 100 bp
* 57285 63653: contig of 6369 bp in length
* 63654 63753: gap of 100 bp
* 63754 105066: contig of 41313 bp in length
* 105067 105166: gap of 100 bp
* 105167 107339: contig of 2173 bp in length
* 107340 107439: gap of 100 bp
* 107440 112075: contig of 4636 bp in length
* 112076 112175: gap of 100 bp
* 112176 115003: contig of 2828 bp in length
* 115004 115103: gap of 100 bp
* 115104 117693: contig of 2590 bp in length.
FEATURES             Location/Qualifiers
     source            1..117693
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="Rp11-13024"
                        /clone_lib="RPC1-11 Human Male BAC"
BASE COUNT      32428 a 26676 c 26081 g 31446 t 1062 others
ORIGIN

alignment_scores:
  Quality:      9.00      Length:      9
  Ratio:        1.000     Gaps:        0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AC016033/rev ..

Align seg 1/1 to reverse of: AC016033 from: 1 to: 117693

      226 ValThrSerLeuGluAsnLeuArgAla 234
      |||||||||||||||||||||||||||
5726  GTCACCTCTCTGGAGATCTCAGGCT 5700

```



OM of: US-09-701-711-2 to: Issued\_Patents\_NA:\* out\_format : pfs

Date: Sep 19, 2002 4:43 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODE=frame+ p2n.model -DEV=xlp  
-O/cgn2_1/USPTO.scopl/US09701711/runat_17092002_141431_28906/app_query.fasta_1.882  
-DB-Issued_Patents_NA -QFMT-fastap -SUFFIX-sepl7oli.rni  
-GAPOP=4.500 -GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000  
-LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000  
-XGAPEXT=60.000 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000  
-YGAPEXT=60.000 -DELOP=6.000 -DELEXT=7.000 -START=1  
-MATRIX=oligo -TRANS=humana40.cdi -LIST=45 -DOCALLIGN=200  
-THR_SCORE=quality -THR_MIN=1 -ALIGN=40 -MODE=LOCAL -OUTFMT=pfs  
-NORM-ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09701711_@CGN1_1_104 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1
```

Search information block:

Query: US-09-701-711-2

Query length: 813

Database: Issued\_Patents\_NA:\*

Database sequences: 383533

Database length: 122816752

Search time (sec): 87.960000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000  
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score\_list:

```
Sequence Strd Orig zScore EScore Len ! Documentation  
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-08-936-165A-220 - 9.00 140.06 4.09 1000  
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-068-880-1 + 8.00 124.71 29.29 604 ! S  
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-068-880-14 + 8.00 120.67 49.20 1026 !  
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-126-109-11 + 8.00 110.16 189.24 4062 !  
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-314-917-1 + 8.00 110.01 193.02 4145 !  
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-265-046-1 + 8.00 110.01 193.02 4145 !  
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-465-522-1 + 8.00 110.01 193.02 4145 !  
/cgn2_6/ptodata/2/ina/PTCTUS.COMB.seq:PCT-US93-11401-1 + 8.00 110.01 193.02 4145  
/cgn2_6/ptodata/2/ina/PTCTUS.COMB.seq:PCT-US95-07849-1 + 8.00 110.01 193.02 4145  
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-08-378-313-20 - 8.00 104.04 415.04 9060 !  
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-103-840A-2 - 8.00 56.82 1.6e+05 4403765  
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-103-840A-1 - 8.00 56.80 1.6e+05 4411529  
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-07-750-080A-6 + 7.00 118.17 67.75 115 !  
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-651-472-6 + 7.00 118.17 67.75 115 ! S  
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-08-358-928-6 + 7.00 118.17 67.75 115 ! S  
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-611-757-96 - 7.00 113.58 122.17 210 !  
/cgn2_6/ptodata/2/ina/PTCTUS.COMB.seq:PCT-US95-05980-96 - 7.00 113.58 122.17 210  
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-143-576-8 - 7.00 113.29 126.72 218 ! S  
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-157-270-13 - 7.00 113.08 130.14 224 !  
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-157-270-11 - 7.00 113.05 130.70 225 !  
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-157-270-10 - 7.00 113.02 131.27 226 !  
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-09-157-270-5 - 7.00 112.95 132.41 228 ! S  
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-840-683-4 + 7.00 112.91 132.98 229 ! S  
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-555-722-4 + 7.00 112.91 132.98 229 ! S  
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-384-301-4 + 7.00 112.91 132.98 229 ! S  
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-157-270-12 - 7.00 112.91 132.98 229 !  
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-485-657A-21 - 7.00 112.59 138.66 239 !  
/cgn2_6/ptodata/2/ina/PTCTUS.COMB.seq:PCT-US95-02303-20 - 7.00 112.59 138.66 239  
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-08-818-112-46 - 7.00 110.20 188.47 327 !  
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-08-818-111-46 - 7.00 110.20 188.47 327 !  
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-056-556-46 - 7.00 110.20 188.47 327 !  
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-467-420A-5 - 7.00 110.03 192.42 334 !  
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-470-110A-5 - 7.00 110.03 192.42 334 !  
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-667-769A-5 - 7.00 110.03 192.42 334 !  
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-940-371-5 - 7.00 110.03 192.42 334 ! S  
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-637-647-5 - 7.00 110.03 192.42 334 ! S  
/cgn2_6/ptodata/2/ina/PTCTUS.COMB.seq:PCT-US95-17082A-5 - 7.00 110.03 192.42 334  
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-441-971-10 + 7.00 109.90 195.81 340 !
```

```
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-08-221-653-10 + 7.00 109.90 195.81 340  
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-08-442-144A-10 + 7.00 109.90 195.81 340  
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-08-441-970-10 + 7.00 109.90 195.81 340  
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-08-818-112-116 - 7.00 108.73 227.33 396  
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-08-818-111-111 - 7.00 108.73 227.33 396  
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-056-556-116 - 7.00 108.73 227.33 396  
/cgn2_6/ptodata/2/ina/6B.COMB.seq:US-09-228-986-33 + 7.00 108.64 230.14 401
```

seq\_name: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:US-08-936-165A-220

seq\_documentation\_block:

```
; Sequence 220, Application US/08936165A  
; Patent No. 6348582  
; GENERAL INFORMATION:  
; APPLICANT: Black, Michael  
; APPLICANT: Burnham, Martin  
; APPLICANT: Hodgson, John  
; APPLICANT: Knowles, David  
; APPLICANT: Lonetto, Michael  
; APPLICANT: Nicholas, Richard  
; APPLICANT: Pratt, Julie  
; APPLICANT: Reichard, Richard  
; APPLICANT: Rosenberg, Martin  
; APPLICANT: Ward, Judith  
; TITLE OF INVENTION: No. 6348582el Prokaryotic Polynucleotides,  
; TITLE OF INVENTION: Polypeptides and Their Uses  
; NUMBER OF SEQUENCES: 534  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SmithKline Beecham Corporation  
; STREET: 709 Swedeland Road  
; CITY: King of Prussia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19406-0939  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/936,165A  
; FILING DATE: 24-SEP-1997  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/027,032  
; FILING DATE: 24-SEP-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gimmi, Edward R  
; REGISTRATION NUMBER: 38,891  
; REFERENCE/DOCKET NUMBER: P50549  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610-270-4478  
; TELEFAX: 610-270-5090  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 220:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1000 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
US-08-936-165A-220
```

alignment\_scores:  
Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-701-711-2 x US-08-936-165A-220/rev ..

Align seg 1/1 to reverse of: US-08-936-165A-220 from: 1 to: 1000

```
717 ValArgproValIlePheIleGluGly 725
|||||
311 GTGGACCGTCATTTTATTGAAGGA 285

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-068-880-1

seq_documentation_block:
; Sequence 1, Application US/09068880B
; Patent No. 6203982
; GENERAL INFORMATION:
; APPLICANT: Nunokawa, Youichi
; APPLICANT: Oikawa, Shinzo
; APPLICANT: Tanaka, Shoji
; TITLE OF INVENTION: Method for Screening Compounds
; TITLE OF INVENTION: Regulating the Expression of Human-Inducible Nitric Oxide
; TITLE OF INVENTION: Synthase
; FILE REFERENCE: SHIM-001
; CURRENT APPLICATION NUMBER: US/09/068,880B
; CURRENT FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: PCT/JP97/03303
; EARLIER FILING DATE: 1997-09-18
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 604
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-068-880-1

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x US-09-068-880-1 ..
Align seg 1/1 to: US-09-068-880-1 from: 1 to: 604

115 GluGlyLeuGlnGluGlyLeuLys 122
|||||
107 GAGGGCTACAGGAGGGTTAAAG 130

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-068-880-14

seq_documentation_block:
; Sequence 14, Application US/09068880B
; Patent No. 6203982
; GENERAL INFORMATION:
; APPLICANT: Nunokawa, Youichi
; APPLICANT: Oikawa, Shinzo
; APPLICANT: Tanaka, Shoji
; TITLE OF INVENTION: Method for Screening Compounds
; TITLE OF INVENTION: Regulating the Expression of Human-Inducible Nitric Oxide
; TITLE OF INVENTION: Synthase
; FILE REFERENCE: SHIM-001
; CURRENT APPLICATION NUMBER: US/09/068,880B
; CURRENT FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: PCT/JP97/03303
; EARLIER FILING DATE: 1997-09-18
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 1026
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-068-880-14

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x US-09-068-880-14 ..
Align seg 1/1 to: US-09-068-880-14 from: 1 to: 1026

115 GluGlyLeuGlnGluGlyLeuLys 122
|||||
2 GAGGGCTACAGGAGGGTTAAAG 25

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-126-109-11

seq_documentation_block:
; Sequence 11, Application US/09126109
; Patent No. 6171856
; GENERAL INFORMATION:
; APPLICANT: Thigpen, Anice
; APPLICANT: Hohmeier, Hans-Ewald
; APPLICANT: Newgard, Christopher B.
; APPLICANT: Unger, Roger H.
; APPLICANT: Shimabukuro, Michio
; APPLICANT: Chen, Guaxun
; APPLICANT: Rhodes, Christopher J.
; APPLICANT: Hugl, Sigrun R.
; APPLICANT: Cousin, Sharon
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
; TITLE OF INVENTION: TO NO-MEDIATED CYTOTOXICITY
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/126,109
; FILING DATE: 30-JUL-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/055,092
; FILING DATE: 30-JUL-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US Unknown
; FILING DATE: 03-MAR-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: McMillian, Nabeela R.
; REGISTRATION NUMBER: P-43,363
; REFERENCE/DOCKET NUMBER: UTSD:560
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4062 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-126-109-11

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

```
alignment_block:
US-09-701-711-2 x US-09-126-109-11 ..
Align seg 1/1 to: US-09-126-109-11 from: 1 to: 4062

115 GluGlyLeuGlnGluGlyLeuLys 122
|||||
3566 GAGGGCCTACAGAGGGGTTAAAG 3589

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-314-917-1

seq_documentation_block:
; Sequence 1, Application US/08314917
; Patent No. 5468630
; GENERAL INFORMATION:
; APPLICANT: Billiar, Timothy R.
; APPLICANT: Nussler, Andreas K.
; APPLICANT: Geller, David A.
; APPLICANT: Simmons, Richard L.
; TITLE OF INVENTION: cDNA Clone for Human Inducible Nitric
; OXIDE SYNTHASE AND PROCESS FOR PREPARING SAME
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold B. Silverman
; STREET: Eckert Seamans Cherin & Mellott
; CITY: Pittsburgh
; STATE: PA
; COUNTRY: USA
; ZIP: 15219
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/314,917
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/981,344
; FILING DATE: 25-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Silverman, Arnold B.
; REGISTRATION NUMBER: 22,614
; REFERENCE/DOCKET NUMBER: 116972
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (412) 566-6000
; TELEFAX: (412) 566-6099
; TELEX: 866172
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4145 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; DESCRIPTION: Human Hepatocyte Inducible Nitric Oxide
; DESCRIPTION: Synthase cDNA Clone
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; TISSUE TYPE: Induced Human Hepatocyte RNA
; IMMEDIATE SOURCE:
; LIBRARY: Lambda zap II cDNA
; CLONE: PHINOS
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: unknown
; MAP POSITION: unknown
; UNITS: unknown
; FEATURE:
; NAME/KEY: CDS

; LOCATION: 207..3668
; IDENTIFICATION METHOD: Experiment
US-08-314-917-1

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x US-08-314-917-1 ..
Align seg 1/1 to: US-08-314-917-1 from: 1 to: 4145

115 GluGlyLeuGlnGluGlyLeuLys 122
|||||
3667 GAGGGCCTACAGAGGGGTTAAAG 3690

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-265-046-1

seq_documentation_block:
; Sequence 1, Application US/08265046
; Patent No. 5658565
; GENERAL INFORMATION:
; APPLICANT: Timothy R. Billiar
; APPLICANT: Edith Tzeng
; APPLICANT: Andreas K. Nussler
; APPLICANT: David A. Geller
; APPLICANT: Richard L. Simmons
; TITLE OF INVENTION: Inducible Nitric Oxide Synthase
; GENE FOR TREATMENT OF DISEASE
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lewis F. Gould, Jr.
; ADDRESS: Eckert Seamans Cherin & Mellott
; STREET: 1700 Market Street, Suite 3232
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/265,046
; FILING DATE: 24-JUN-1994
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Gould, Lewis F. Jr.
; REGISTRATION NUMBER: 25,057
; REFERENCE/DOCKET NUMBER: 119130
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 575-6020
; TELEFAX: (215) 575-6015
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4145 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; DESCRIPTION: Human Hepatocyte Inducible Nitric Oxide
; DESCRIPTION: Synthase cDNA Clone
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; TISSUE TYPE: Induced Human Hepatocyte RNA
; IMMEDIATE SOURCE:
; LIBRARY: Lambda zap II cDNA
```

CLONE: PHINOS  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: unknown  
MAP POSITION: unknown  
UNITS: unknown  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 207..3668  
IDENTIFICATION METHOD: Experiment  
US-08-265-046-1

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-701-711-2 x US-08-265-046-1 ..

Align seg 1/1 to: US-08-265-046-1 from: 1 to: 4145

115 GluGlyLeuGlnGluGlyLeuLys 122  
|||||  
3667 GAGGCGCTACAGGAGGGTTAAAG 3690

seq\_name: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:US-08-465-522-1

seq\_documentation\_block:  
Sequence 1, Application US/08465522  
Patent No. 5882908  
GENERAL INFORMATION:  
APPLICANT: Billiar, Timothy R.  
APPLICANT: Nussler, Andreas K.  
APPLICANT: Geller, David A.  
APPLICANT: Simmons, Richard L.  
TITLE OF INVENTION: cDNA Clone for Human Inducible Nitric  
TITLE OF INVENTION: Oxide Synthase And Process for Preparing Same  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lewis F. Gould, Jr.  
ADDRESSEE: Eckert Seamans Cherin & Mellott  
STREET: 1700 Market St. Suite 3232  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/465,522  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Gould, Jr., Lewis F.  
REGISTRATION NUMBER: 25,057  
REFERENCE/DOCKET NUMBER: 116972-6  
TELEPHONE: (215) 575-6020  
TELEFAX: (215) 575-6015  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:  
LENGTH: 4145 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
DESCRIPTION: Human Hepatocyte Inducible Nitric Oxide  
DESCRIPTION: Synthase cDNA Clone

HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
TISSUE TYPE: Induced Human Hepatocyte RNA  
IMMEDIATE SOURCE:  
LIBRARY: Lambda Zap II cDNA  
CLONE: PHINOS  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: unknown  
MAP POSITION: unknown  
UNITS: unknown  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 207..3668  
IDENTIFICATION METHOD: Experiment  
US-08-465-522-1

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-701-711-2 x US-08-465-522-1 ..

Align seg 1/1 to: US-08-465-522-1 from: 1 to: 4145

115 GluGlyLeuGlnGluGlyLeuLys 122  
|||||  
3667 GAGGCGCTACAGGAGGGTTAAAG 3690

seq\_name: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:PCT-US93-11401-1

seq\_documentation\_block:  
Sequence 1, Application PCTUS9311401  
GENERAL INFORMATION:  
APPLICANT: Billiar, Timothy R.  
APPLICANT: Nussler, Andreas K.  
APPLICANT: Geller, David A.  
APPLICANT: Simmons, Richard L.  
TITLE OF INVENTION: cDNA Clone for Human Inducible Nitric  
TITLE OF INVENTION: Oxide Synthase And Process for Preparing Same  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold B. Silverman  
ADDRESSEE: Eckert Seamans Cherin & Mellott  
STREET: 600 Grant Street, 42nd Floor  
CITY: Pittsburgh  
STATE: PA  
COUNTRY: USA  
ZIP: 15219  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/11401  
FILING DATE: 25-NOV-1992  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/981,344  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Silverman, Arnold B.  
REGISTRATION NUMBER: 22,614  
REFERENCE/DOCKET NUMBER: 116972  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (412) 566-6000  
TELEFAX: (412) 566-6099  
TELEX: 866172  
INFORMATION FOR SEQ ID NO: 1:

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4145 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; TISSUE TYPE: Induced Human Hepatocyte RNA
; IMMEDIATE SOURCE:
; LIBRARY: Lambda zap II cDNA
; CLONE: PHINOS
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: unknown
; MAP POSITION: unknown
; UNITS: unknown
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 207..3668
; IDENTIFICATION METHOD: Experiment
; PCT-US93-11401-1

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x PCT-US93-11401-1 ..
Align seg 1/1 to: PCT-US93-11401-1 from: 1 to: 4145

115 GluGlyLeuGlnGluGlyLeuLys 122
|||||
3667 GAGGGCCTACAGGAGGGTTAAAG 3690

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq: PCT-US95-07849-1

seq_documentation_block:
; Sequence 1, Application PC/TUS9507849
; GENERAL INFORMATION:
; APPLICANT: University of Pittsburgh of the Commonwealth System of Higher
; APPLICANT: Education
; TITLE OF INVENTION: Inducible Nitric Oxide Synthase
; TITLE OF INVENTION: Gene for Treatment of Disease
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lewis F. Gould, Jr.
; ADDRESSEE: Eckert Seamans Cherin & Mellott
; STREET: 1700 Market Street, Suite 3232
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07849
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gould, Lewis F. Jr.
; REGISTRATION NUMBER: 25,057
; REFERENCE/DOCKET NUMBER: 119130-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 575-6020
; TELEFAX: (215) 575-6015
; TELEX:
```

```
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4145 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; DESCRIPTION: Human Hepatocyte Inducible Nitric Oxide
; DESCRIPTION: Synthase cDNA Clone
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; TISSUE TYPE: Induced Human Hepatocyte RNA
; IMMEDIATE SOURCE:
; LIBRARY: Lambda zap II cDNA
; CLONE: PHINOS
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: unknown
; MAP POSITION: unknown
; UNITS: unknown
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 207..3668
; IDENTIFICATION METHOD: Experiment
; PCT-US95-07849-1

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x PCT-US95-07849-1 ..
Align seg 1/1 to: PCT-US95-07849-1 from: 1 to: 4145

115 GluGlyLeuGlnGluGlyLeuLys 122
|||||
3667 GAGGGCCTACAGGAGGGTTAAAG 3690

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq: US-08-378-313-20

seq_documentation_block:
; Sequence 20, Application US/08378313
; Patent No. 6207881
; GENERAL INFORMATION:
; APPLICANT: THEOLOGIS, ATHANASIOS
; APPLICANT: SATO, TAKAHIDO
; TITLE OF INVENTION: CONTROL OF FRUIT RIPENING THROUGH
; TITLE OF INVENTION: GENETIC CONTROL OF ACC SYNTHASE SYNTHESIS
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/378.313
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/862,493
; FILING DATE: 02-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
```

```
;
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 29190-20002.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 856-5600
; TELEFAX: (415) 494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9060 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(2704..2880, 2968..3099, 3183..3344, 3810
; LOCATION: ..4376, 4463..4903)
; US-08-378-313-20

alignment_scores:
    Quality: 8.00      Length: 8
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x US-08-378-313-20/rev ..

Align seg 1/1 to reverse of: US-08-378-313-20 from: 1 to: 9060

500 ThrLysTYrAspAsnLysAsnIle 507
|||||
5398 ACAAAATATGACACACAAAACATC 5375

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-2

seq_documentation_block:
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

alignment_scores:
    Quality: 8.00      Length: 8
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x US-09-103-840A-2/rev ..

Align seg 1/1 to reverse of: US-09-103-840A-2 from: 1 to: 4403765

685 ThrThrLeuGlyGluValValGly 692
|||||
```

```
3909845ACCACGCTGGGGAGGTGTCGGA 3909822
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-1

seq_documentation_block:
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

alignment_scores:
    Quality: 8.00      Length: 8
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x US-09-103-840A-1/rev ..

Align seg 1/1 to reverse of: US-09-103-840A-1 from: 1 to: 4411529

685 ThrThrLeuGlyGluValValGly 692
|||||
3916067ACCACGCTGGGGAGGTGTCGGA 3916044

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-750-080A-6

seq_documentation_block:
; Sequence 6, Application US/07750080A
; Patent No. 5445953
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: DIRECT MOLECULAR CLONING OF A MODIFIED
; TITLE OF INVENTION: EUKARYOTIC CYTOPLASMIC DNA VIRUS GENOME
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/750,080A
; FILING DATE: 19910826
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/106 IMMU
```

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 115 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
IMMEDIATE SOURCE:  
CLONE: PA0 (Fig. 4.3)  
US-07-750-080A-6

alignment\_scores:  
Quality: 7.00 Length: 7  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-701-711-2 x US-07-750-080A-6 ..

Align seg 1/1 to: US-07-750-080A-6 from: 1 to: 115

409 ArgProValProAsnSerPro 415  
|||||  
90 AGGCCAGTACCAATTCGCC 110

seq\_name: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:US-08-651-472-6

seq\_documentation\_block:  
Sequence 6, Application US/08651472  
Patent No. 6103244  
GENERAL INFORMATION:

APPLICANT: DORNER, Friedrich  
APPLICANT: SCHEIFLINGER, Friedrich  
APPLICANT: FALKNER, Falko Gunter  
APPLICANT: PFELEIDERER, Michael  
TITLE OF INVENTION: DIRECT MOLECULAR CLONING OF CHIMERIC  
TITLE OF INVENTION: VIRUSES CONTAINING HUMAN IMMUNODEFICIENCY VIRUS TYPE 1  
NUMBER OF SEQUENCES: 95  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109

COMPUTER READABLE FORM:  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/651,472  
FILING DATE:

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/914,738  
FILING DATE: 20-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/750,080  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/166/IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399

TELEX: 904136  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 115 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Other nucleic acid;  
IMMEDIATE SOURCE:  
CLONE: PA0  
US-08-651-472-6

alignment\_scores:  
Quality: 7.00 Length: 7  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-701-711-2 x US-08-651-472-6 ..

Align seg 1/1 to: US-08-651-472-6 from: 1 to: 115

409 ArgProValProAsnSerPro 415  
|||||  
90 AGGCCAGTACCAATTCGCC 110

seq\_name: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:US-08-358-928-6

seq\_documentation\_block:  
Sequence 6, Application US/08358928  
Patent No. 6265183  
GENERAL INFORMATION:

APPLICANT: DORNER, Friedrich  
APPLICANT: SCHEIFLINGER, Friedrich  
APPLICANT: FALKNER, Falko Gunter  
APPLICANT: PFELEIDERER, Michael  
TITLE OF INVENTION: DIRECT MOLECULAR CLONING OF CHIMERIC  
TITLE OF INVENTION: VIRUSES CONTAINING HUMAN IMMUNODEFICIENCY VIRUS TYPE 1  
NUMBER OF SEQUENCES: 95  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/358,928  
FILING DATE:

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/914,738  
FILING DATE: 20-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/750,080  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/166/IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 6:

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 115 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Other nucleic acid;
; DESCRIPTION: Synthetic DNA oligonucleotide
; IMMEDIATE SOURCE:
; CLONE: pA0
; US-08-358-928-6

alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x US-08-358-928-6 ..
Align seg 1/1 to: US-08-358-928-6 from: 1 to: 115
409 ArgProValProAsnSerPro 415
|||||
90 AGGCCAGTACCCATTGCCCC 110

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-611-757-96.
seq_documentation_block:
; Sequence 96, Application US/08611757
; Patent No. 5859230
; GENERAL INFORMATION:
; APPLICANT: Kim, Jungsuh P.
; APPLICANT: Reyes, Gregory R.
; APPLICANT: Wages, John
; APPLICANT: Zhang-keck, Zhen-yang
; APPLICANT: Young, Lavonne
; TITLE OF INVENTION: No. 5859230-A/No. 5859230-B/No. 5859230-C/No. 5859230-D/No. 58
; TITLE OF INVENTION: Agents and Molecular Cloning Thereof
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/611,757
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/246,985
; FILING DATE: 20-MAY-1994
; APPLICATION NUMBER: US 025,396
; FILING DATE: 24-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/922,493
; FILING DATE: 30-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-0201
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 96:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 115 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: SCH Clone SU6-2
```

```
; LENGTH: 210 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: SCH Clone SU6-2
; US-08-611-757-96

alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x US-08-611-757-96/rev ..
Align seg 1/1 to reverse of: US-08-611-757-96 from: 1 to: 210
787 SerIleSerTyrAlaLysPro 793
|||||
91 TCGATTTCCTACGCCAAACCC 71

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-05980-96
seq_documentation_block:
; Sequence 96, Application PC/TUS9505980
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT:
; TITLE OF INVENTION: Non-A/Non-B/Non-C/Non-D/Non-E Hepatitis
; TITLE OF INVENTION: Agents and Molecular Cloning Thereof
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/05980
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/246,986
; FILING DATE: 20-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-0201.49
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 96:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 210 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: SCH Clone SU6-2
```



PCT-US95-05980-96

alignment\_scores:  
Quality: 7.00 Length: 7  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-701-711-2 x PCT-US95-05980-96/rev ..

Align seg 1/1 to reverse of: PCT-US95-05980-96 from: 1 to: 210

787.SerIleSerTyrAlaLysPro 793  
|||||  
91 TCGATTTCCTAGCCCAACCC 71

seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-143-576-8

seq\_documentation\_block:  
; Sequence 8, Application US/08143576  
; Patent No. 5643761

; GENERAL INFORMATION:  
; APPLICANT: Fisher, Paul B.  
; APPLICANT: Jiang, Hongping  
; TITLE OF INVENTION: METHOD FOR GENERATING A SUBSTRACTED  
; TITLE OF INVENTION: CDNA LIBRARY AND USES OF THE GENERATED LIBRARY  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRES:  
; ADDRESSEE: John P. White, c/o Cooper & Dunham  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10112

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/143,576  
; FILING DATE: 25-OCT-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P.  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 43563/JPW/AKC  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 977-9550  
; TELEFAX: (212) 664-0525  
; TELEX: 422523 COOP UI  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
; HYPOTHEICAL: NO  
; ANTI-SENSE: NO  
US-08-143-576-8

alignment\_scores:  
Quality: 7.00 Length: 7  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-701-711-2 x US-08-143-576-8/rev ..

Align seg 1/1 to reverse of: US-08-143-576-8 from: 1 to: 218

472 ArgGluValTyrSerLeuGly 478  
|||||  
216 CGAGAGGTTTACAGTCTAGGT 196

seq\_name: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:US-09-157-270-13

seq\_documentation\_block:  
; Sequence 13, Application US/09157270  
; Patent No. 6306401

; GENERAL INFORMATION:  
; APPLICANT: Johnson, Eric S.  
; APPLICANT: Pham, Thuy D.  
; TITLE OF INVENTION: Viral Detection System  
; FILE REFERENCE: D6161  
; CURRENT APPLICATION NUMBER: US/09/157,270  
; CURRENT FILING DATE: 1998-09-18  
; EARLIER APPLICATION NUMBER: US 60/061,287  
; EARLIER FILING DATE: 1997-10-07  
; NUMBER OF SEQ ID NOS: 26  
; SEQ ID NO 13  
; LENGTH: 224  
; TYPE: DNA  
; ORGANISM: unknown

; FEATURE:  
; OTHER INFORMATION: RT-PCR product from egg albumin of commercial chicken from  
; OTHER INFORMATION: randomly chosen grocery store # 205.  
US-09-157-270-13

alignment\_scores:  
Quality: 7.00 Length: 7  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-701-711-2 x US-09-157-270-13/rev ..

Align seg 1/1 to reverse of: US-09-157-270-13 from: 1 to: 224

289 GluLeuGluAlaLeuLys 295  
|||||  
189 GAGTTAGAACGCGTGTAAAG 169

seq\_name: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:US-09-157-270-11

seq\_documentation\_block:  
; Sequence 11, Application US/09157270  
; Patent No. 6306401

; GENERAL INFORMATION:  
; APPLICANT: Johnson, Eric S.  
; APPLICANT: Pham, Thuy D.  
; TITLE OF INVENTION: Viral Detection System  
; FILE REFERENCE: D6161  
; CURRENT APPLICATION NUMBER: US/09/157,270  
; CURRENT FILING DATE: 1998-09-18  
; EARLIER APPLICATION NUMBER: US 60/061,287  
; EARLIER FILING DATE: 1997-10-07  
; NUMBER OF SEQ ID NOS: 26  
; SEQ ID NO 11  
; LENGTH: 225  
; TYPE: DNA  
; ORGANISM: Single Comb White Leghorn chicken

; FEATURE:  
; OTHER INFORMATION: RT-PCR product from egg albumin of stock 70 chicken.  
US-09-157-270-11

alignment\_scores:  
Quality: 7.00 Length: 7  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

```
alignment_block:
US-09-701-711-2 x US-09-157-270-11/rev ..
Align seg 1/1 to reverse of: US-09-157-270-11 from: 1 to: 225

289 GluLeuGluAlaLeuLeuLys 295
|||||
191 GAGTTAGAGCGCTGTTAAAG 171

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-157-270-10
seq_documentation_block:
; Sequence 10, Application US/09157270
; Patent No. 6306401
; GENERAL INFORMATION:
; APPLICANT: Johnson, Eric S.
; APPLICANT: Pham, Thuy D.
; TITLE OF INVENTION: Viral Detection System
; FILE REFERENCE: D6161
; CURRENT APPLICATION NUMBER: US/09/157,270
; CURRENT FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: US 60/061,287
; EARLIER FILING DATE: 1997-10-07
; NUMBER OF SEQ ID NOS: 26
; SEQ ID NO 10
; LENGTH: 226
; TYPE: DNA
; ORGANISM: Single Comb White Leghorn chicken
; FEATURE:
; OTHER INFORMATION: RT-PCR product from egg albumin of stock 6F chicken.
US-09-157-270-10

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x US-09-157-270-10/rev ..
Align seg 1/1 to reverse of: US-09-157-270-10 from: 1 to: 226

289 GluLeuGluAlaLeuLeuLys 295
|||||
191 GAGTTAGAGCGCTGTTAAAG 171

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-157-270-5
seq_documentation_block:
; Sequence 5, Application US/09157270
; Patent No. 6306401
; GENERAL INFORMATION:
; APPLICANT: Johnson, Eric S.
; APPLICANT: Pham, Thuy D.
; TITLE OF INVENTION: Viral Detection System
; FILE REFERENCE: D6161
; CURRENT APPLICATION NUMBER: US/09/157,270
; CURRENT FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: US 60/061,287
; EARLIER FILING DATE: 1997-10-07
; NUMBER OF SEQ ID NOS: 26
; SEQ ID NO 5
; LENGTH: 228
; TYPE: DNA
; ORGANISM: Avian leukosis/sarcoma virus of the family Retroviridae
; FEATURE:
; OTHER INFORMATION: Subgroup E of the avian leukosis/sarcoma virus.
US-09-157-270-5

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

```
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x US-09-157-270-5/rev ..
Align seg 1/1 to reverse of: US-09-157-270-5 from: 1 to: 228

289 GluLeuGluAlaLeuLeuLys 295
|||||
194 GAGTTAGAGCGCTGTTAAAG 174

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-840-683-4
seq_documentation_block:
; Sequence 4, Application US/08840683
; Patent No. 5821051
; GENERAL INFORMATION:
; APPLICANT: Androphy, Elliot J.
; APPLICANT: Chen, Jason J.
; TITLE OF INVENTION: E6-BINDING PROTEINS
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/840,683
; FILING DATE: 29-APR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/273,059
; FILING DATE: 08-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: NEP-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 229 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-840-683-4

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x US-08-840-683-4 ..
Align seg 1/1 to: US-08-840-683-4 from: 1 to: 229

52 ValLeuProPheArgLeuGly 58
|||||
58 GTTTTACCATTAGCTGGCT 78

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-555-722-4
```

```
seq_documentation_block:
; Sequence 4, Application US/08555722
; Patent No. 5989804
; GENERAL INFORMATION:
; APPLICANT: Androphy, Elliot J.
; APPLICANT: Chen, Jason J.
; TITLE OF INVENTION: EG-BINDING PROTEINS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/555,722
; FILING DATE: 14-NOV-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: NEP-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 229 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-08-555-722-4

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x US-08-555-722-4 ..
Align seg 1/1 to: US-08-555-722-4 from: 1 to: 229

52 ValLeuPropheArgLeuGly 58
|||||
58 GTTTACCATTTAGGCTGGGT 78

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-384-301-4

seq_documentation_block:
; Sequence 4, Application US/09384301
; Patent No. 6296853
; GENERAL INFORMATION:
; APPLICANT: Androphy, Elliot J.
; APPLICANT: Chen, Jason J.
; TITLE OF INVENTION: EG-BINDING PROTEINS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
```

```
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/384,301
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/555,722
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: NEP-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 229 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-09-384-301-4

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x US-09-384-301-4 ..
Align seg 1/1 to: US-09-384-301-4 from: 1 to: 229

52 ValLeuPropheArgLeuGly 58
|||||
58 GTTTACCATTTAGGCTGGGT 78

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-157-270-12

seq_documentation_block:
; Sequence 12, Application US/09157270
; Patent No. 6306401
; GENERAL INFORMATION:
; APPLICANT: Johnson, Eric S.
; APPLICANT: Pham, Thuy D.
; TITLE OF INVENTION: Viral Detection System
; FILE REFERENCE: D6161
; CURRENT APPLICATION NUMBER: US/09/157,270
; CURRENT FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: US 60/061,287
; EARLIER FILING DATE: 1997-10-07
; NUMBER OF SEQ ID NOS: 26
; SEQ ID NO 12
; LENGTH: 229
; TYPE: DNA
; ORGANISM: Single Comb white Leghorn chicken
; FEATURE:
; OTHER INFORMATION: RT-PCR product from egg albumin of stock 10Q chicken.
US-09-157-270-12

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
```

```
US-09-701-711-2 x US-09-157-270-12/rev ..
Align seg 1/1 to reverse of: US-09-157-270-12 from: 1 to: 229
289 GluLeuGluAlaLeuLeuLys 295
|||||
194 GAGTTAGAAGCGCTGTTAAAG 174

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-485-657A-21
seq_documentation_block:
; Sequence 21, Application US/08485657A
; Patent No. 5942389
; GENERAL INFORMATION:
; APPLICANT: Kirschling, Deborah J
; APPLICANT: Gudkov, Andrei
; APPLICANT: Roninson, Igor B
; TITLE OF INVENTION: Genes and Genetic Elements Associated
; TITLE OF INVENTION: with Sensitivity to Cisplatin
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive, 32nd Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/485,657A
; APPLICATION NUMBER: 07-JUN-1995
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5942389man, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 93,354-N
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; TELEX:
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 239 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; ANTI-SENSE: YES
US-08-485-657A-21

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x US-08-485-657A-21/rev ..
Align seg 1/1 to reverse of: US-08-485-657A-21 from: 1 to: 239
289 GluLeuGluAlaLeuLeuLys 295
|||||
153 GAACGGAGGCGCTCTGAAA 133

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-02303-20
seq_documentation_block:
; Sequence 20, Application PC/TUS9502303
```

```
; GENERAL INFORMATION:
; APPLICANT: Genes and Genetic Elements Associated
; TITLE OF INVENTION: with Sensitivity to Cisplatin
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION NUMBER: PCT/US95/02303
; APPLICATION NUMBER: PCT/US95/02303
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 239 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; ANTI-SENSE: YES
PCT-US95-02303-20

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x PCT-US95-02303-20/rev ..
Align seg 1/1 to reverse of: PCT-US95-02303-20 from: 1 to: 239
289 GluLeuGluAlaLeuLeuLys 295
|||||
153 GAACGGAGGCGCTCTGAAA 133

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-818-112-46
seq_documentation_block:
; Sequence 46, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
```

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 327 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-818-112-46

alignment\_scores:  
Quality: 7.00 Length: 7  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-701-711-2 x US-08-818-112-46/rev ..

Align seg 1/1 to reverse of: US-08-818-112-46 from: 1 to: 327

259 ArgillePheValGlulIleSer 265  
|||||

226 CGAATATTGTCGAGATCTCG 206

seq\_name: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:US-08-818-111-46

seq\_documentation\_block:

; Sequence 46, Application US/08818111  
; Patent No. 6338852

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonia  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Twardzik, Daniel R.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF  
; NUMBER OF SEQUENCES: 148

; CORRESPONDENCE ADDRESS: 148 TUBERCULOSIS

; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/818,111

; FILING DATE: 13-MAR-1997

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

; REGISTRATION NUMBER: 31,392

; REFERENCE/DOCKET NUMBER: 210121.417C6

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 46:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 327 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-818-111-46

alignment\_scores:  
Quality: 7.00 Length: 7  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-701-711-2 x US-08-818-111-46/rev ..

Align seg 1/1 to reverse of: US-08-818-111-46 from: 1 to: 327

259 ArgillePheValGlulIleSer 265

|||||

226 CGAATATTGTCGAGATCTCG 206

seq\_name: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:US-09-056-556-46

seq\_documentation\_block:

; Sequence 46, Application US/09056556  
; Patent No. 6350456

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Dillon, Davin C.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND

; NUMBER OF SEQUENCES: 241

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/056,556

; FILING DATE: 07-APR-1998

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

; REGISTRATION NUMBER: 31,392

; REFERENCE/DOCKET NUMBER: 210121.457

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 46:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 327 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-09-056-556-46

alignment\_scores:  
Quality: 7.00 Length: 7  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-701-711-2 x US-09-056-556-46/rev ..

Align seg 1/1 to reverse of: US-09-056-556-46 from: 1 to: 327

259 ArgillePheValGlulIleSer 265

|||||

226 CGAATATTGTCGAGATCTCG 206

seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-467-420A-5

TRE

seq\_documentation\_block:  
; Sequence 5, Application US/08467420A  
; Patent No. 5683892  
; GENERAL INFORMATION:  
; APPLICANT: Ames, Robert S.  
; APPLICANT: Appelbaum, Edward R.  
; APPLICANT: Chaiken, Irwin M.  
; APPLICANT: Cook, Richard M.  
; APPLICANT: Gross, Mitchell S.  
; APPLICANT: Holmes, Stephen D.  
; APPLICANT: McMillan, Lynette J.  
; APPLICANT: Theisen, Timothy W.  
; TITLE OF INVENTION: Recombinant IL5 Antagonists Useful in  
; TREATMENT OF IL5 Mediated Disorders  
; NUMBER OF SEQUENCES: 74  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SmithKline Beecham Corp./Corporate  
; ADDRESSEE: Intellectual Property  
; STREET: P. O. Box 1539-UW2220  
; CITY: King of Prussia  
; STATE: Pennsylvania  
; COUNTRY: USA  
; ZIP: 19406-0939  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/467,420A  
; FILING DATE:  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/363131  
; FILING DATE: 23-DEC-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sutton, Jeffrey A.  
; REGISTRATION NUMBER: 34,028  
; REFERENCE/DOCKET NUMBER: P50282  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610 270-5024  
; TELEFAX: 610 270-5090  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 334 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 1..334  
; OTHER INFORMATION: /note= "First base corresponds to  
; OTHER INFORMATION: Kabat position 24"  
US-08-467-420A-5

alignment\_scores:  
Quality: 7.00 Length: 7  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-701-711-2 x US-08-467-420A-5/rev ..

Align seg 1/1 to reverse of: US-08-467-420A-5 from: 1 to: 334

669 LeuGlyProArgSerGlnAla 675  
|||||  
311 CTGGCCCCAGAGTCAGGCC 291

seq\_name: /cgn2\_6/ptodata/2/1na/5A\_COMB.seq:US-08-470-110A-5

seq\_documentation\_block:  
; Sequence 5, Application US/08470110A  
; Patent No. 5693323  
; GENERAL INFORMATION:  
; APPLICANT: Ames, Robert S.  
; APPLICANT: Appelbaum, Edward R.  
; APPLICANT: Chaiken, Irwin M.  
; APPLICANT: Cook, Richard M.  
; APPLICANT: Gross, Mitchell S.  
; APPLICANT: Holmes, Stephen D.  
; APPLICANT: McMillan, Lynette J.  
; APPLICANT: Theisen, Timothy W.  
; TITLE OF INVENTION: Recombinant IL5 Antagonists Useful in  
; TREATMENT OF IL5 Mediated Disorders  
; NUMBER OF SEQUENCES: 74  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SmithKline Beecham Corp./Corporate  
; ADDRESSEE: Intellectual Property  
; STREET: P. O. Box 1539-UW2220  
; CITY: King of Prussia  
; STATE: Pennsylvania  
; COUNTRY: USA  
; ZIP: 19406-0939  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/470,110A  
; FILING DATE:  
; CLASSIFICATION: 426  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/363131  
; FILING DATE: 23-DEC-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sutton, Jeffrey A.  
; REGISTRATION NUMBER: 34,028  
; REFERENCE/DOCKET NUMBER: P50282  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610 270-5024  
; TELEFAX: 610 270-5090  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 334 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 1..334  
; OTHER INFORMATION: /note= "First base corresponds to  
; OTHER INFORMATION: Kabat position 24"  
US-08-470-110A-5

alignment\_scores:  
Quality: 7.00 Length: 7  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-701-711-2 x US-08-470-110A-5/rev ..

Align seg 1/1 to reverse of: US-08-470-110A-5 from: 1 to: 334

669 LeuGlyProArgSerGlnAla 675  
|||||  
311 CTGGCCCCAGAGTCAGGCC 291

seq\_name: /cgn2\_6/ptodata/2/1na/5A\_COMB.seq:US-08-667-769A-5

```

seq_documentation_block:
; Sequence 5, Application US/08667769A
; Patent No. 5783184
; GENERAL INFORMATION:
; APPLICANT: Ames, Robert S.
; APPLICANT: Appelbaum, Edward R.
; APPLICANT: Chaiken, Irwin M.
; APPLICANT: Cook, Richard M.
; APPLICANT: Gross, Mitchell S.
; APPLICANT: Holmes, Stephen D.
; APPLICANT: McMillan, Lynette J.
; APPLICANT: Theisen, Timothy W.
; TITLE OF INVENTION: Recombinant IL5 Antagonists Useful in
; TITLE OF INVENTION: Treatment of IL5 Mediated Disorders
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corp./Corporate
; STREET: P.O. Box 1539-UW2220
; CITY: King of Prussia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/667,769A
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/17082
; FILING DATE: 22-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/470110
; FILING DATE: 06-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/467420
; FILING DATE: 06-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/363131
; FILING DATE: 23-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Sutton, Jeffrey A.
; REGISTRATION NUMBER: 34,028
; REFERENCE/DOCKET NUMBER: P50503
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5024
; TELEFAX: 610-270-5090
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 334 base pairs
; TYPE: nucleic acid
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..334
; OTHER INFORMATION: /note= "First base corresponds to
; US-08-667-769A-5

```

```

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-701-711-2 x US-08-667-769A-5/rev

```

```

Align seq 1/1 to reverse of: US-08-667-769A-5 from: 1 to: 334

669 LeuGlyProArgSerGlnAla 675
|||||
311 CTTGGCCCCCAGAAAGTCAAGCC 291

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-940-371-5

seq_documentation_block:
; Sequence 5, Application US/08940371
; Patent No. 5851525
; GENERAL INFORMATION:
; APPLICANT: Ames, Robert S.
; APPLICANT: Appelbaum, Edward R.
; APPLICANT: Chaiken, Irwin M.
; APPLICANT: Cook, Richard M.
; APPLICANT: Gross, Mitchell S.
; APPLICANT: Holmes, Stephen D.
; APPLICANT: McMillan, Lynette J.
; APPLICANT: Theisen, Timothy W.
; TITLE OF INVENTION: Recombinant IL5 Antagonists Useful in
; TITLE OF INVENTION: Treatment of IL5 Mediated Disorders
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corp./Corporate
; ADDRESS: Intellectual Property
; STREET: P. O. Box 1539-UW2220
; CITY: King of Prussia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/940,371
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,110
; FILING DATE:
; APPLICATION NUMBER: US 08/363131
; FILING DATE: 23-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Sutton, Jeffrey A.
; REGISTRATION NUMBER: 34,028
; REFERENCE/DOCKET NUMBER: P50282
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610 270-5024
; TELEFAX: 610 270-5090
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 334 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..334
; OTHER INFORMATION: /note= "First base corresponds to
; US-08-940-371-5

```

```

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:

```

```
US-09-701-711-2 x US-08-940-371-5/rev ..
Align seg 1/1 to reverse of: US-08-940-371-5 from: 1 to: 334

669 LeuGlyProArgSerGlnAla 675
|||||
311 CTGGCCCCAGAGTCAAGCC 291

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-637-647-5

seq documentation block:
; Sequence 5, Application US/08637647
; Patent No. 6129913
; GENERAL INFORMATION:
; APPLICANT: Ames, Robert S.
; APPLICANT: Appelbaum, Edward R.
; APPLICANT: Chaiken, Irwin M.
; APPLICANT: Cook, Richard M.
; APPLICANT: Gross, Mitchell S.
; APPLICANT: McMillan, Lynette J.
; APPLICANT: Theisen, Timothy W.
; TITLE OF INVENTION: Recombinant IL5 Antagonists Useful in
; TREATMENT OF IL5 Mediated Disorders
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corp./Corporate
; ADDRESSEE: Intellectual Property
; STREET: P. O. Box 1539-UM2220
; CITY: King of Prussia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/637,647
; FILING DATE: 19-JUN-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/363,131
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Sutton, Jeffrey A.
; REGISTRATION NUMBER: 34,028
; REFERENCE/DOCKET NUMBER: P50282
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610 270-5024
; TELEFAX: 610 270-5090
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 334 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1..334
; OTHER INFORMATION: /note= "First base corresponds to
; Kabat position 24"
US-08-637-647-5

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
seq documentation block:
; Sequence 5, Application PC/TUS9517082A
; GENERAL INFORMATION:
; APPLICANT: Ames, Robert S.
; APPLICANT: Appelbaum, Edward R.
; APPLICANT: Chaiken, Irwin M.
; APPLICANT: Cook, Richard M.
; APPLICANT: Gross, Mitchell S.
; APPLICANT: Holmes, Stephen D.
; APPLICANT: McMillan, Lynette J.
; APPLICANT: Theisen, Timothy W.
; TITLE OF INVENTION: Recombinant IL5 Antagonists Useful in
; TREATMENT OF IL5 Mediated Disorders
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corp./Corporate
; STREET: P.O. Box 1539-UM2220
; CITY: King of Prussia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/17082A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/470110
; FILING DATE: 06-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/467420
; FILING DATE: 06-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/363131
; FILING DATE: 23-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Sutton, Jeffrey A.
; REGISTRATION NUMBER: 34,028
; REFERENCE/DOCKET NUMBER: P50282-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5024
; TELEFAX: 610-270-5090
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 334 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1..334
; OTHER INFORMATION: /note= "First base corresponds to
; Kabat position 24"
PCT-US95-17082A-5

alignment_scores:
Quality: 7.00 Length: 7
```



Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-09-701-711-2 x PCT-US95-17082A-5/rev ..  
Align seg 1/1 to reverse of: PCT-US95-17082A-5 from: 1 to: 334  
seq\_name: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:US-08-441-971-10  
669 LeuGlyProArgSerGlnAla 675  
311 CTGTGCCCCGAGAGTCAAGCC 291  
seq\_documentation\_block:  
; Sequence 10, Application US/08441971  
; Patent No. 6071693  
; GENERAL INFORMATION:  
; APPLICANT: Tai-An Cha  
; TITLE OF INVENTION: HCV GENOMIC SEQUENCES FOR  
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 147  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
; STREET: 600 Atlantic Avenue  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02210  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 5.25 inch  
; COMPUTER: IBM compatible  
; OPERATING SYSTEM: MS-DOS Version 3.3  
; SOFTWARE: WordPerfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/441,971  
; FILING DATE: 16-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/221,653  
; FILING DATE:  
; FILING DATE: 07/697,326  
; FILING DATE:  
; FILING DATE: 8 May 1991  
; NAME: Janluk, Anthony J.  
; REGISTRATION NUMBER: 29,809  
; REFERENCE/DOCKET NUMBER: C0772/7000  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 720-3500  
; TELEFAX: (617) 720-2441  
; TELEX: EZEKIEL  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 340 nucleotides  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: ns5gh6  
US-08-441-971-10  
alignment\_scores:  
Quality: 7.00 Length: 7  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-09-701-711-2 x US-08-441-971-10 ..

Align seg 1/1 to: US-08-441-971-10 from: 1 to: 340  
669 LeuGlyProArgSerGlnAla 675  
58 CTGTGCCCCCGAAGCCAGGCA 78  
seq\_name: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:US-08-221-653-10  
seq\_documentation\_block:  
; Sequence 10, Application US/08221653  
; Patent No. 6190864  
; GENERAL INFORMATION:  
; APPLICANT: Tai-An Cha  
; TITLE OF INVENTION: HCV GENOMIC SEQUENCES FOR  
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 147  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
; STREET: 600 Atlantic Avenue  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02210  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 5.25 inch  
; COMPUTER: IBM compatible  
; OPERATING SYSTEM: MS-DOS Version 3.3  
; SOFTWARE: WordPerfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/221,653  
; FILING DATE:  
; FILING DATE: 07/697,326  
; FILING DATE:  
; FILING DATE: 8 May 1991  
; NAME: Janluk, Anthony J.  
; REGISTRATION NUMBER: 29,809  
; REFERENCE/DOCKET NUMBER: C0772/7000  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 720-3500  
; TELEFAX: (617) 720-2441  
; TELEX: EZEKIEL  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 340 nucleotides  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: ns5gh6  
US-08-221-653-10  
alignment\_scores:  
Quality: 7.00 Length: 7  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-09-701-711-2 x US-08-221-653-10 ..  
Align seg 1/1 to: US-08-221-653-10 from: 1 to: 340  
669 LeuGlyProArgSerGlnAla 675  
58 CTGTGCCCCCGAAGCCAGGCA 78  
seq\_name: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:US-08-442-144A-10

```
seq_documentation_block:
; Sequence 10, Application US/08442144A
; Patent No. 6214583
; GENERAL INFORMATION:
; APPLICANT: Tai-An Cha
; APPLICANT: Eileen Beall
; APPLICANT: Bruce Irvine
; APPLICANT: Janice Kolberg
; APPLICANT: Michael S. Urdea
; TITLE OF INVENTION: HCV GENOMIC SEQUENCES FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 Inch
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows NT
; SOFTWARE: Microsoft Word 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/442,144A
; FILING DATE: MAY 16, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/221,653
; FILING DATE: APRIL 1, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Doreen Yanko Trujillo
; REGISTRATION NUMBER: 35,719
; REFERENCE/DOCKET NUMBER: CHIR-0121
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; TELEX:
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 340 Nucleotides
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE: DNA
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: ns5gh6
; US-08-442-144A-10
```

```
alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

```
alignment_block:
```

```
US-09-701-711-2 x US-08-442-144A-10 ..
```

```
Align seg 1/1 to: US-08-442-144A-10 from: 1 to: 340
```

```
669 LeuGlyProArgSerGlnAla 675
|||||
58 CTTGGCCCCCGAAGCCAGGCA 78
```

OM of: US-09-701-711-2 to: EST:\* out\_format : pfs

Date: Sep 19, 2002 3:44 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framet p2n.model -DEV=xlp  
-O=/cgn2.1/USPTO.spool/US09701711/runat\_17092002.141431\_28874/app\_query.fasta\_1.882  
-DB=EST -QFWT=fastap -SUFFIX=sepl7oli.rst -GAPOP=4.500  
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -GAPOP=60.000 -XGAPEXT=60.000  
-GAPOP=6.000 -GAPEXT=7.000 -GAPOP=60.000 -XGAPEXT=60.000  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=quality  
-THR\_MIN=1 -ALIGN=40 -MODE=LOCAL -OUTFWT=pfs -NORM=ext  
-HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09701711 -CGN1\_1=5751 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-701-711-2  
Query length: 813  
Database: EST:\*  
Database sequences: 13736207  
Database length: -1841457050  
Search time (sec): 1940.450000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000  
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score\_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_est1:BM209113	-	9.00	155.68	20.04	180
gb_est1:AI528822	-	9.00	153.13	27.77	238
gb_est1:BM010871	-	9.00	151.69	33.43	279
gb_est1:BM123030	-	9.00	148.57	49.87	393
gb_est1:BM840995	-	9.00	148.45	50.61	398
gb_est1:AV625127	-	9.00	147.04	60.69	465
gb_gss:AAQ66784	+	9.00	146.90	61.75	472
gb_gss:AAQ837806	+	9.00	146.71	63.28	482
gb_gss:AAQ836677	-	9.00	145.25	76.34	566
gb_gss:BH025710	+	9.00	145.25	76.34	566
gb_gss:AAQ509377	+	9.00	144.84	80.44	592
gb_gss:AAQ290748	+	9.00	144.55	83.47	611
gb_est1:BM173688	+	9.00	144.32	86.02	627
gb_gss:AAZ391700	+	9.00	144.20	87.31	635
gb_est1:BM1998945	+	9.00	144.00	89.56	649
gb_est1:BM177050	+	9.00	143.92	90.52	655
gb_gss:AAQ80404D	+	9.00	140.84	134.41	919
gb_est1:BM397547	+	9.00	137.18	214.93	1374
gb_est1:BM420056	+	9.00	135.81	256.17	1597
gb_est1:AAW796076	+	8.00	140.24	145.08	116
gb_gss:AAZ286261	-	8.00	140.01	149.47	119
gb_gss:AAQ809005	+	8.00	138.80	174.67	136
gb_est1:BM983453	+	8.00	135.15	278.78	203
gb_est1:AV257342	-	8.00	135.02	283.60	206
gb_est1:BM128835	-	8.00	134.93	286.82	208
gb_est1:R70943	+	8.00	134.63	298.11	215
gb_est1:BF377641	+	8.00	134.54	301.35	217
gb_est1:BM1493603	+	8.00	133.90	327.44	233
gb_est1:AA729757	-	8.00	133.48	345.56	244
gb_est1:BM417742	-	8.00	133.15	360.48	253
gb_est1:BM579201	+	8.00	132.97	368.81	258
gb_est1:AAZ42705	+	8.00	132.83	375.49	262
gb_est1:AA1866213	+	8.00	132.62	385.55	268
gb_est1:AA1583432	+	8.00	132.42	395.64	274
gb_est1:BM537706	+	8.00	132.32	400.70	277
gb_gss:AAZ375019	+	8.00	132.32	400.70	277
gb_est1:BM2436	+	8.00	131.81	427.84	293
gb_est1:AV205177	-	8.00	131.75	431.26	295

gb\_est1:AU100452 + 8.00 131.60 439.80 300 ! AU100452 AU100452 Sugano Hom  
gb\_est1:AI528822 - 8.00 131.54 443.22 302 ! AI528822 ms15c10.x1 Stratage  
gb\_est1:BM428988 - 8.00 131.51 444.94 303 ! BF428988 WHE1712.G05.N102S.W  
gb\_est1:AA776501 + 8.00 131.42 450.08 306 ! AA776501 ah11f11.s1 Gessler  
gb\_est1:BM0311 + 8.00 131.42 450.08 306 ! BM0311 yr41e09.s1 Soares fet  
gb\_gss:BM596414 - 8.00 131.18 463.85 304 ! BM596414 BOGEA797F BOGE Brns  
gb\_est1:BM905044 - 8.00 131.01 474.21 320 ! BM905044 TaLr1136607F TaLr1

seq\_name: gb\_est2:BM209113

seq\_documentation\_block:

LOCUS BM209113 180 bp mRNA linear EST 31-JAN-2002  
DEFINITION C0640C12-3 NIA Mouse Trophoblast Stem Cell cDNA Library (Long) Mus  
musculus cDNA clone C0640C12 3', mRNA sequence.

ACCESSION BM209113

VERSION BM209113.1 GI:17765496

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 180)

AUTHORS Piao,Y., Kargul,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Luo,A.,

Tanaka,T., Kunath,T., Rossant,J. and Ko,M.S.H.

TITLE Systematic Analyses of NIA Mouse Trophoblast Stem Cell cDNA Library

JOURNAL (Long)

COMMENT Unpublished (2001)

Contact: Dawood B. Dudekula

Laboratory of Genetics

National Institute on Aging/National Institutes of Health

333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA

Email: cdna@lgsun.grc.nia.nih.gov

Plate: C0640 row: C column: 12

Seq primer: -21M13 Forward

High quality sequence stop: 180

POLYA=Yes.

Location/Qualifiers

1..180

/organism="Mus musculus"

/strain="B5/EGFP transgenic ICR mice"

/db\_xref="niaEST:C0640C12-3"

/db\_xref="taxon:10090"

/clone="C0640C12"

/cldna\_lib="NIA Mouse Trophoblast Stem Cell cDNA Library

(Long)"

/tissue\_type="Trophoblast stem cell"

/dev\_stage="3.5-dpc"

/lab\_host="DH10B"

/notes="Vector: pSPOT1 (Invitrogen); Site\_1: SalI; Site\_2:

NotI; Mouse cDNA project by the Laboratory of Genetics,

National Institute on Aging (NIA), Intramural Research

Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is

a long-transcript enriched cDNA library (Ref. Genome Res.

11: 1553-1558 (2001). [PMID: 11544199]]. Total RNAs were

obtained from Dr. Janet Rossant and Tilo Kunath (Samuel

Lunenfeld Research Institute, Canada). Double-stranded

cDNAs were synthesized with an Oligo(dT) primer

[Invitrogen]:

5'-pGACTAGTCTAGATCCGAGCGCCCTTTTCTTTT-3' from

4 ug of total RNA, treated with T4 DNA polymerase, and

purified by ethanol-precipitation. The cDNAs were ligated

to Lone-linker LL-Sal4, purified by phenol/chloroform, and

separated from free linkers by Centricon 100. Then, the

cDNAs were amplified by long-range high fidelity PCR using

Ex Taq polymerase (Takara) with a primer Sal4-S. The

products were purified by phenol/chloroform and Centricon

100. The cDNAs were digested with SalI and NotI enzymes

and cloned into SalI/NotI site of pSPOT1 plasmid vector.

The DH10B E. coli host was transformed with the ligation

mixture by the standard chemical method. The average

insert size is about 2.6 kb. The library was constructed

by Yulan Piao (NIA)."

49 a 38 c 50 g 43 t

BASE COUNT

## ORIGIN

alignment\_scores:  
 Quality: 9.00 Length: 9  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-701-711-2 x BM209113/rev ..

Align seg 1/1 to reverse of: BM209113 from: 1 to: 180

705 LeuProLeuProPheLysGlyAspTrp 713

|||||  
 53 CTGCCCTACCCCTCAAGGAGATTGG 27

seq\_name: gb\_estl:A1886635

## seq\_documentation\_block:

LOCUS A1886635 238 bp mRNA linear EST 07-MAR-2000  
 DEFINITION wm60h01.x1 NCI\_CGAP\_Ut2 Homo sapiens cDNA clone IMAGE:2440369 3',  
 mRNA sequence.

ACCESSION A1886635

VERSION A1886635.1 GI:5591799

KEYWORDS EST.

SOURCE human.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

1 (bases 1 to 238)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: [cgapsb@mail.nih.gov](mailto:cgapsb@mail.nih.gov)

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Insert Length: 1944 Std Error: 0.00

Seq primer: -40UP from Gibco

High quality sequence stop: 217.

## FEATURES

source

1..238  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2440369"  
 /clone\_lib="NCI\_CGAP\_Ut2"  
 /tissue\_type="moderately-differentiated endometrial  
 adenocarcinoma, 3 pooled tumors"  
 /lab\_host="DH10B"  
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: SalI;  
 Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 1.85 kb. Life Technologies catalog #:  
 11539-012"

## BASE COUNT

ORIGIN 89 a 40 c 25 g 84 t

## alignment\_scores:

Quality: 9.00 Length: 9  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-701-711-2 x A1886635 ..

Align seg 1/1 to: A1886635 from: 1 to: 238

703 LeuLeuProLeuProPheLysGly 711

|||||  
 188 CTTATACCTCCACTTCCTTTAAAGGA 214

seq\_name: gb\_estl:BB010871

## seq\_documentation\_block:

LOCUS BB010871 279 bp mRNA linear EST 22-JUN-2000  
 DEFINITION BB010871 RIKEN full-length enriched, 10 day neonate skin Mus  
 musculus cDNA clone 4732498B18 3', mRNA sequence.

ACCESSION BB010871

VERSION BB010871.1 GI:8131228

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 279)

## AUTHORS

Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci  
 P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,  
 Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,  
 Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,  
 Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.,  
 Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,  
 Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata  
 Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.,  
 Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomimaga,N., Toya  
 T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.,  
 Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino  
 M., Muramatsu,M. and Hayashizaki,Y.  
 RIKEN Mouse ESTs (Konno,H., et al.)  
 Unpublished (2000)

## TITLE

JOURNAL

COMMENT

Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: [genome-res@gsc.riken.go.jp](mailto:genome-res@gsc.riken.go.jp),  
 URL:<http://genome.gsc.riken.go.jp/>  
 Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki  
 N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
 Thermostabilization and thermoactivation of thermolabile enzymes by  
 trehalose and its application for the synthesis of full length  
 cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
 Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,  
 Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki  
 Y. and Hayashizaki,Y.  
 Automated filtration-based high-throughput plasmid preparation  
 system. Genome Res. 9 (5), 463-470 (1999)  
 Carninci,P. and Hayashizaki,Y.  
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
 19-44 (1999)  
 Please visit our web site (<http://genome.rtc.riken.go.jp>) for  
 further details.

## FEATURES

source

1..279  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="4732498B18"  
 /clone\_lib="RIKEN full-length enriched, 10 day neonate  
 skin"  
 /sex="mixed"  
 /tissue\_type="skin"  
 /dev\_stage="10 days neonate"  
 /lab\_host="DH10B"

/note="Site 1: SalI; Site 2: BamHI; cDNA library was  
 prepared and sequenced in Mouse Genome Encyclopedia  
 Project of Genome Exploration Research Group in Riken



derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 107 a 94 c 100 g 97 t  
ORIGIN

## alignment\_scores:

Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-701-711-2 x BE840995/rev ..  
Align seg 1/1 to reverse of: BE840995 from: 1 to: 398

683 GlnGlnThrThrLeuGlyGluValVal 691  
|||||  
42 CAGCAACGACCCCTTGCTGAAGTCGTC 16

seq\_name: gb\_est1:AV625127

## seq\_documentation\_block:

LOCUS AV625127 465 bp mRNA linear EST 15-DEC-2000  
DEFINITION AV625127 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas reinhardtii cDNA clone LC088c03\_r 5', mRNA sequence.

ACCESSION AV625127  
VERSION AV625127.1 GI:10774304  
KEYWORDS EST.

SOURCE Chlamydomonas reinhardtii.

## ORGANISM

Chlamydomonas reinhardtii.  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
Chlamydomonadaceae; Chlamydomonas.

## REFERENCE

AUTHORS Asamizu, E., Miura, K., Kuchio, K., Inoue, Y., Fukuzawa, H., Ohyama, K., Nakamura, Y., and Tabata, S.

TITLE Generation of expressed sequence tags from low-CO2 and high-CO2 adapted cells of Chlamydomonas reinhardtii

JOURNAL DNA Res. 7 (5), 305-307 (2000)

MEDLINE 20539644

COMMENT Contact: Erika Asamizu

The First Laboratory for Plant Gene Research

Kazusa DNA Research Institute

Yana 1532-3, Kisarazu, Chiba 292-0812, Japan

Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

Location/Qualifiers

## FEATURES

source

1..465  
/organism="Chlamydomonas reinhardtii"  
/strain="C9"  
/db\_xref="taxon:3055"  
/clone="LC088c03\_r"  
/clone\_lib="Chlamydomonas reinhardtii 5% to 0.04% CO2"  
/note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2: XhoI; The cDNA library was constructed from cells cultured in a carbon stress acclimatized condition in which carbon dioxide concentration in the bubbling gas was changed from 5% to 0.04%"

BASE COUNT 118 a 143 c 104 g 100 t  
ORIGIN

## alignment\_scores:

Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-701-711-2 x AV625127 ..

Align seg 1/1 to: AV625127 from: 1 to: 465

49 LeuGlnSerValLeuProPheArgLeu 57  
|||||  
177 TTGCAATCAGTGTCCCTTCGTTG 203

seq\_name: gb\_gss:AQ667844

## seq\_documentation\_block:

LOCUS AQ667844 472 bp DNA linear GSS 24-JUN-1999  
DEFINITION HS\_5383\_AL\_G02\_MR RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate-959 Col-3 Row-M, DNA sequence.

ACCESSION AQ667844

VERSION AQ667844.1 GI:5200590

KEYWORDS GSS.

SOURCE human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 472)

## AUTHORS

Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Kellar, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L.

## TITLE

Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

MEDLINE 99380589

## COMMENT

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm)

or from Resear h Genetics (info@resgen.com). BAC end Web Server:

http://www.htsc.washington.edu

Plate: 959 row: M column: 3

Seq primer: 17

Class: BAC ends

High quality sequence stop: 472.

## FEATURES

source

Location/Qualifiers  
1..472  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Plate-959 Col-3 Row-M"  
/clone\_lib="RPCI-11 Human Male BAC Library"  
/sex="male"  
/note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at EcoRI sites"  
BASE COUNT 123 a 118 c 89 g 141 t  
ORIGIN

## alignment\_scores:

Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-701-711-2 x AQ667844 ..

Align seg 1/1 to: AQ667844 from: 1 to: 472

585 GlyTrpAsnTyrSerSerLeuAspArg 593

|||||

212 GGATGGAACATATAGCTCTCTGGATAGA 238

seq\_name: gb\_gss:AQ837806

seq\_documentation\_block:

LOCUS AQ837806 482 bp DNA linear GSS 30-AUG-1999  
DEFINITION HS\_521\_A2\_F12\_SP6E RPCI-11 Human Male BAC Library Homo sapiens  
genomic clone Plate-1097 Col-24 Row-K, DNA sequence.

ACCESSION AQ837806  
VERSION AQ837806.1 GI:5807680  
KEYWORDS GSS.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 482)  
Mahairas G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,  
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and  
Hood, L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
MEDLINE 99380589

COMMENT High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
(pieterdejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm)  
or from Resear h Genetics (info@resgen.com). BAC end Web Server:  
http://www.htsc.washington.edu  
Plate: 1097 row: K column: 24  
Seq primer: SP6  
Class: BAC ends  
High quality sequence stop: 482.

#### FEATURES

source

Location/Qualifiers

1..482  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Plate-1097 Col-24 Row-K"  
/clone\_lib="RPCI-11 Human Male BAC Library"  
/sex="male"

/note="Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
Male blood DNA was isolated from one randomly chosen donor  
and partially digested with a combination of EcoRI and  
EcoRI Methylase. Size selected DNA was cloned into the  
pBACE3.6 vector at EcoRI sites"

BASE COUNT 185 a 64 c 86 g 139 t 8 others

ORIGIN

#### alignment\_scores:

Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

#### alignment\_block:

US-09-701-711-2 x AQ837806/rev

Align seg 1/1 to reverse of: AQ837806 from: 1 to: 482

786 LeuSerIleSerTyrAlaIysProLeu 794

|||||  
172 TTGACGATCTCATGCTAAACCAATTA 146

seq\_name: gb\_gss:AQ836677

seq\_documentation\_block:

LOCUS AQ836677 566 bp DNA linear GSS 20-FEB-2001  
DEFINITION 2M0131D07R Mouse 10kb plasmid UUGC1M library Mus musculus genomic

clone UUGC2M0131D07 R, DNA sequence.

ACCESSION A2836677  
VERSION A2836677.1 GI:13006585  
KEYWORDS GSS.  
SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 566)  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.  
and Wright, D., Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)

CONTACT: Robert B. Weiss  
University of Utah Genome Center

University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA

Tel: 801 585 5606  
Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00  
Plate: 0131 row: D column: 07

Seq primer: CACACAGGAACACGCTATGACC

Class: plasmid ends

High quality sequence stop: 566.

#### FEATURES

source

Location/Qualifiers

1..566  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC2M0131D07"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adaptor DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pWD42 (gi14732114|gb|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptor mouse DNA was annealed to  
adaptor vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT 161 a 128 c 135 g 142 t

ORIGIN

#### alignment\_scores:

Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

#### alignment\_block:

US-09-701-711-2 x A2836677/rev

Align seg 1/1 to reverse of: A2836677 from: 1 to: 566

61 ValSerGluAsnGlnLeuAlaAspGly 69

|||||

180 GTTTCAGAAAACAGCTAGCAGATGGT 154

seq\_name: gb\_gss:BH025710

seq\_documentation\_block:

LOCUS BH025710 566 bp DNA linear GSS 17-JUL-2001  
 DEFINITION RPCI-24-259D20.TV RPCI-24 Mus musculus genomic clone RPCI-24-259D20

, DNA sequence.

ACCESSION BH025710

VERSION BH025710.1 GI:14789174

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 566)

Zhao,S., Nierman,W., Malek,J., Shvartsman,S., Akinret,B., Levins,M.,

Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,

Russell,D., de Jong,P. and Fraser,C.M.

Mouse BAC End Sequences from Library RPCI-24

Unpublished (1999)

Other\_GSSs: RPCI-24-259D20.TJ

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC

library availability, please contact Pieter de Jong

(pdejong@mail.cho.org). Clones may be purchased from BACPAC

Resources (<http://www.tigr.org/bacpac/orderingframe.htm>). BAC end

page: [http://www.tigr.org/tadb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tadb/bac_ends/mouse/bac_end_intro.html)

Plate: 259 row: D column: 20

Seq primer: T7

Class: BAC ends.

Location/Qualifiers

source

1..566

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="RPCI-24-259D20"

/clone\_lib="RPCI-24"

/sex="Male"

/cell\_type="Spleen/Brain"

/note="Vector: pTARBAC1; Site\_1: BamH1; Site\_2: BamH1;

RPCI-24 Mouse BAC Library produced by Pieter de Jong. The

library was cloned in the pTARBAC1 cloning vector at the

BamH1 sites using MboI partially digested male C57BL/6J

DNA."

BASE COUNT 125 a 116 c 181 g 144 t

ORIGIN

alignment\_scores:

Quality: 9.00 Length: 9

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-701-711-2 x BH025710 ..

Align seg 1/1 to: BH025710 from: 1 to: 566

367 GlnAspGluValLeuArgGluMet 375

|||||

177 CAGATGAGGCTTAAAGGAGACAGATG 203

seq\_name: gb\_gss:AQ509377

seq\_documentation\_block:

LOCUS AQ509377

592 bp DNA linear GSS 04-MAY-1999

DEFINITION

nbxb0096K15r CUGI Rice BAC Library Oryza sativa genomic clone  
 nbxb0096K15r, DNA sequence.

ACCESSION

AQ509377

VERSION

AQ509377.1 GI:4733460

KEYWORDS

GSS.

SOURCE

Oryza sativa.

ORGANISM

Oryza sativa

Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartioideae; Oryzeae; Oryza.

1 (bases 1 to 592)

Wing,R.A. and Dean,R.A.

A BAC End Sequencing Framework to Sequence the Rice Genome

Unpublished (1998)

Contact: Wing RA

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: rwing@clemson.edu

Seq primer: GGAACAGCTATCACCATG

Class: BAC ends

High quality sequence stop: 276.

Location/Qualifiers

source

1..592

/organism="Oryza sativa"

/strain="Japonica"

/cultivar="Nipponbare"

/db\_xref="taxon:4530"

/clone="nbxb0096K15r"

/clone\_lib="CUGI Rice BAC Library"

/tissue\_type="Leaf"

/lab\_host="E. coli DH10B"

/note="Vector: pBelobAC11; Site\_1: HindIII; Site\_2:

HindIII; Rice is one of two most popular grains in the

world. Half of the world population especially those

inhabiting highly populated areas of the humid tropics

and subtropics, rely on rice as their primary source of

carbohydrate. Monocotyledonous rice is a diploid plant

(2n=24) with a haploid genome equivalent of 431 Mbp

(Arumuganathan and Earle, 1991). The relatively small

genome of rice, three times larger than that of

Arabidopsis, makes it suitable for genomic studies. In

order to facilitate positional cloning, physical mapping

and genome sequencing of rice, we have constructed a BAC

library from Oryza sativa, Nipponbare variety. The

library contains 36,864 clones with an average insert size

of 128.5 Kb providing 10.9 haploid genome equivalents. The

deep coverage allows the isolation a particular sequence

with a probability of 99.9 %. Two high density filters,

each containing 18,432 clones (doubly spotted), represent

the whole library for colony screening."

BASE COUNT 172 a 111 c 118 g 189 t

ORIGIN

alignment\_scores:

Quality: 9.00 Length: 9

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-701-711-2 x AQ509377/rev ..

Align seg 1/1 to reverse of: AQ509377 from: 1 to: 592

507 IleSerAsnTyrValLeuAspSerTyr 515

|||||

435 ATACAAATTTATGACTAGTAGCTAT 409

seq\_name: gb\_gss:AQ290748





```

KEYWORDS
SOURCE      house mouse.
ORGANISM    Mus musculus
REFERENCE   1 (bases 1 to 635)
AUTHORS     Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D., Weiss,R.
TITLE       Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
JOURNAL     Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0154 row: C column: 06
            Seq primer: CGTTGTAAACGACGCCAGT
            Class: plasmid ends
            High quality sequence stop: 635.

FEATURES             source
1. .635
   /organism="Mus musculus"
   /strain="C57BL/6J"
   /db_xref="taxon:10090"
   /clone="UUGCLM0154C06"
   /clone_lib="Mouse 10kb plasmid UUGCLM library"
   /sex="Male"
   /lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
   /note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (g14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT      245 a 117 c 147 g 126 t
ORIGIN

alignment_scores:
  Quality:      9.00      Length:      9
  Ratio:        1.000     Gaps:      0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AZ391700/rev ..

Align seg 1/1 to reverse of: AZ391700 from: 1 to: 635

701 SerGluLeuLeuLeuProLeuProPhe 709
|||||
503 TCAGAACTGATATGCCCTCCCTTC 477

seq_name: gb_est2:BI998945

GSS.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 635)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0154 row: C column: 06
Seq primer: CGTTGTAAACGACGCCAGT
Class: plasmid ends
High quality sequence stop: 635.

FEATURES             source
1. .635
   /organism="Mus musculus"
   /strain="C57BL/6J"
   /db_xref="taxon:10090"
   /clone="UUGCLM0154C06"
   /clone_lib="Mouse 10kb plasmid UUGCLM library"
   /sex="Male"
   /lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
   /note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (g14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT      245 a 117 c 147 g 126 t
ORIGIN

alignment_scores:
  Quality:      9.00      Length:      9
  Ratio:        1.000     Gaps:      0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AZ391700/rev ..

Align seg 1/1 to reverse of: AZ391700 from: 1 to: 635

701 SerGluLeuLeuLeuProLeuProPhe 709
|||||
503 TCAGAACTGATATGCCCTCCCTTC 477

seq_name: gb_est2:BI998945

seq_documentation_block:
LOCUS      BI998945          649 bp      mRNA      linear      EST 25-OCT-2001
DEFINITION Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION  BI998945
VERSION     BI998945.1 GI:16433719
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
ORGANISM    Chlamydomonas reinhardtii.
REFERENCE   1 (bases 1 to 649)
AUTHORS     Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre
            ,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.
            Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants. Project: 1031
            Unpublished (2001)
            Contact: Charles Hauser
            DCMB Box 91000
            Duke University
            Durham, NC 27708-1000
            Tel: 919 613 8159
            Fax: 919 613 8177
            Email: chauser@duke.edu.
            Location/Qualifiers
            1. .649
              /organism="Chlamydomonas reinhardtii"
              /strain="CC-1690 wild type mt+ 21gr"
              /db_xref="taxon:3055"
              /clone_lib="C. reinhardtii CC-1690, Stress II (normalized
              ), Lambda Zap II"
              /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
              XhoI; Stress condition II library, constructed by John
              Davies and Jeffrey McDermott, combines cDNAs from CC-1690
              cells grown to mid-log phase in TAP (NH4+ - containing)
              and shifted to TAP - NO3- (24hrs); H2 production
              conditions (0, 12hr, 24hr) see Mellis et al., (2000) Plant
              Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP +
              sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr).
              POLYA mRNA was purified from each sample, pooled and cDNA
              synthesized. The cDNA was directionally cloned into lambda
              Zap II (Stratagene) in the EcoRI (5') and XhoI (3')
              sites. pBluescript II SK- plasmids were excised from the
              lambda Zap clones by superinfection with EXAssist
              (Stratagene) phage. The library was normalized using
              method 4 described in Bonaldo et al., (1996) Genome
              Research 6: 791-806."
              Research 6: 791-806."
BASE COUNT      162 a 207 c 145 g 134 t 1 others
ORIGIN

alignment_scores:
  Quality:      9.00      Length:      9
  Ratio:        1.000     Gaps:      0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x BI998945 ..

Align seg 1/1 to: BI998945 from: 1 to: 649

49 LeuGlnSerValLeuProPheArgLeu 57
|||||
183 TTGCAATCAGTGTCCCTTCGTTG 209

seq_name: gb_est2:BG177050

seq_documentation_block:
LOCUS      BG177050          655 bp      mRNA      linear      EST 06-FEB-2001
DEFINITION 602313868F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4419520 5',
            mRNA sequence.

```

**ACCESSION** BG177050  
**VERSION** BG177050.1 GI:12683753  
**KEYWORDS** EST.  
**SOURCE** human.  
**ORGANISM** Homo sapiens  
**REFERENCE** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
**AUTHORS** 1 (bases 1 to 655)  
**TITLE** NIH-MGC http://mgs.nci.nih.gov/.  
**JOURNAL** National Institutes of Health, Mammalian Gene Collection (MGC)  
**COMMENT** Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: Louis Staudt, M.D., Ph.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
 Plate: L1AM10155 row: h column: 17  
 High quality sequence stop: 471.  
**FEATURES**  
 source  
 Location/Qualifiers  
 1..655  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4419520"  
 /clone\_lib="NIH-MGC\_85"  
 /tissue\_type="lymphoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: lymph; Vector: pCMV-Sport6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.867 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH\_MGC Library."  
**BASE COUNT** 184 a 141 c 112 g 218 t  
**ORIGIN**  
 alignment\_scores  
 Quality: 9.00 Length: 9  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block  
 US-09-701-711-2 x BG177050 ..  
 Align seg 1/1 to: BG177050 from: 1 to: 655  
 387 LysileGinLeuSerArgAlaArgLeu 395  
 ||||||||||||||||||||||||||||  
 256 AGATTACGCTCTCTCGTGCAGCTTG 282  
 seq\_name: gb\_gss:CNS04ADV  
 seq\_documentation\_block:  
 LOCUS CNS04ADV 919 bp DNA linear GSS 21-MAY-2000  
 DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone 095G04 of library G from Tetraodon nigroviridis, genomic survey sequence.  
 ACCESSION AL281740  
 VERSION AL281740.1 GI:8020066  
 KEYWORDS GSS: genome survey sequence.  
 SOURCE Tetraodon nigroviridis.  
 ORGANISM Tetraodon nigroviridis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.  
**REFERENCE** 1 (bases 1 to 919)  
**AUTHORS** Roest-Crolius.H., Jaillon.O., Dasilva.C., Fizesma.C., Fisher.C., Bouneau.L., Billault.A., Quetier.F., Saurin.W., Bernot.A. and Weissenbach.J.

**TITLE** Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis  
**JOURNAL** Unpublished  
**REFERENCE** 2 (bases 1 to 919)  
**AUTHORS** Roest-Crolius.H., Jaillon.O., Dasilva.C., Bouneau.L., Fisher.C., Bernot.A., Fizesma.C., Wincker.P., Brottier.P., Quetier.F., Saurin.W. and Weissenbach.J.  
**TITLE** Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence  
**JOURNAL** Unpublished  
**REFERENCE** 3 (bases 1 to 919)  
**AUTHORS** Genoscope.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
**COMMENT** This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/tetraodon.  
**FEATURES**  
 source  
 Location/Qualifiers  
 1..919  
 /organism="Tetraodon nigroviridis"  
 /db\_xref="taxon:99883"  
 /clone="095G04"  
 /clone\_lib="G"  
 /note="Genoscope sequence ID : C0BG095BD02LPI-end : T7"  
**BASE COUNT** 232 a 212 c 241 g 229 t  
**ORIGIN**  
 alignment\_scores  
 Quality: 9.00 Length: 9  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block  
 US-09-701-711-2 x CNS04ADV ..  
 Align seg 1/1 to: CNS04ADV from: 1 to: 919  
 121 LeuLysAsnAlaGlyLeuAlaValGly 129  
 ||||||||||||||||||||||||||||  
 536 CTAAGAAGACGTGGGTAGCTGTCGGG 562  
 seq\_name: gb\_est2:BG397547  
 seq\_documentation\_block:  
 LOCUS BG397547 1374 bp mRNA linear EST 12-MAR-2001  
 DEFINITION 602439379f1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:456586 5', mRNA sequence.  
 ACCESSION BG397547  
 VERSION BG397547.1 GI:13290995  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
**REFERENCE** 1 (bases 1 to 1374)  
**AUTHORS** NIH-MGC http://mgs.nci.nih.gov/.  
**TITLE** National Institutes of Health, Mammalian Gene Collection (MGC)  
**JOURNAL** Unpublished (1999)  
**COMMENT** Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
 CDNA Library Preparation: Ling Hong/Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
 Plate: L1CMI282 row: n column: 19  
 High quality sequence start: 33  
 High quality sequence stop: 261.  
**FEATURES**  
 Location/Qualifiers

```

source
1. .1374
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:456586"
/tissue_type="NIH_MGC_48"
/lab_host="DH10B (phage-resistant)"
/note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; CDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
Note: this is a NIH_MGC Library." 1 others
BASE COUNT      299 a 347 c 425 g 302 t
ORIGIN

alignment_scores:
Quality:      9.00      Length:      9
Ratio:        1.000     Gaps:      0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x BG420056 ..
Align seg 1/1 to: BG420056 from: 1 to: 1597

437 GlyTyrSerGlnSerGlyGlyValThr 445
|||||
1348 GGTACTCACAATCGGTGGTGTGACA 1374

seq_name: gb_est1:AW796076

seq_documentation_block:
LOCUS      AW796076                116 bp      mRNA      linear      EST 16-MAY-2000
DEFINITION MR2-UM0025-300300-102-a09 UM0025 Homo sapiens cDNA, mRNA sequence.
ACCESSION  AW796076
VERSION    AW796076.1 GI:7847946
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 116)
AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
            Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
            Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-MR2-UM0025-300
300-102-a09&t3-2000-03-30&t4-1)
Seq primer: puc 18 forward
High quality sequence stop: 115.
Location/Qualifiers
            1..116
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_lib="UM0025"
            /note="Organ: uterus; Vector: puc18; Site_1: SmaI; Site_2:
            SmaI; A mini-library was made by cloning products derived
            from ORESTES PCR (U.S. Letters Patent application No. 196
            716 - Ludwig Institute for Cancer Research) profiles
            into the pUC 18 vector. Reverse transcription of tissue
            mRNA and cDNA amplification were performed under low
            stringency conditions."

FEATURES
source
1. .1597
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4591882"
/tissue_type="renal cell adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: kidney; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'

```

```

1. .1374
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:456586"
/tissue_type="NIH_MGC_48"
/lab_host="DH10B (phage-resistant)"
/note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; CDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
Note: this is a NIH_MGC Library." 1 others
BASE COUNT      299 a 347 c 425 g 302 t
ORIGIN

alignment_scores:
Quality:      9.00      Length:      9
Ratio:        1.000     Gaps:      0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x BG397547/rev ..
Align seg 1/1 to reverse of: BG397547 from: 1 to: 1374

461 HisValAsnAlaSerPheSerArgSer 469
|||||
175 CATGTCATCAAGCTTTTCCCGATCC 149

seq_name: gb_est2:BG420056

seq_documentation_block:
LOCUS      BG420056                1597 bp      mRNA      linear      EST 14-MAR-2001
DEFINITION 602453743F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4591882 5',
            mRNA sequence.
ACCESSION  BG420056
VERSION    BG420056.1 GI:13326562
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 1597)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgabs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTp
            CDNA Library Preparation: Ling Hong/Rubin Laboratory
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LCM31329 row: f column: 11
            High quality sequence stop: 3.
            Location/Qualifiers
            1..1597
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:4591882"
            /tissue_type="renal cell adenocarcinoma"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: kidney; Vector: pOTB7; Site_1: XhoI; Site_2:
            EcoRI; cDNA made by oligo-dT priming. Directionally
            cloned into EcoRI/XhoI sites using the following 5'

```

BASE COUNT 22 a 41 c 36 g 17 t  
ORIGIN

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-701-711-2 x AW796076 ..

Align seg 1/1 to: AW796076 from: 1 to: 116

662 ArgGlyTyrAspGlnSerSerLeu 669

|||||  
34 CGCGGGTATGACCATGCCAGCCTA 57

seq\_name: gb\_gss:AZ286261

seq\_documentation\_block:

LOCUS AZ286261 119 bp DNA linear GSS 27-JUL-2000  
DEFINITION RPCI-23-442122.TJ RPCI-23 Mus musculus genomic clone RPCI-23-442122  
, DNA sequence.

ACCESSION AZ286261  
VERSION A2286261.1 GI:9528137

KEYWORDS GSS.  
SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 119)  
AUTHORS Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret

,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.  
and Fraser,C.M.  
Mouse BAC End Sequences from Library RPCI-23

TITLE Mouse BAC End Sequences from Library RPCI-23

JOURNAL

COMMENT

Other GSSs: RPCI-23-442122.TV

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)

or from Resea ch Genetics (info@resgen.com). BAC end page:

http://www.tigr.org/tdb/bac.ends/mouse/bac\_end\_intro.html

Plate: 442 row: I column: 22

Seq primer: SP6

Class: BAC ends.

FEATURES

source

1. .119

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="RPCI-23-442122"

/clone\_lib="RPCI-23"

/sex="Female"

/lab\_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site\_1:  
EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or

brain genomic DNA was isolated and partially digested

with a combination of EcoRI and EcoRI Methylase. Size

selected DNA was cloned into the pBACE3.6 vector at the

EcoRI sites. The ligation products were transformed into

DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 18 a 30 c 39 g 32 t

ORIGIN

alignment\_scores:

Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-701-711-2 x AZ286261/rev ..

Align seg 1/1 to reverse of: AZ286261 from: 1 to: 119

177 AlaGlyLysProAlaArgVal 184

|||||

77 GCAGAGGCGACCGCCAGGCTA 54

seq\_name: gb\_gss:AZ909005

seq\_documentation\_block:

LOCUS AZ909005 136 bp DNA linear GSS 05-MAR-2001  
DEFINITION RPCI-24-220018.TJ RPCI-24 Mus musculus genomic clone RPCI-24-220018  
, DNA sequence.

ACCESSION AZ909005

VERSION A2909005.1 GI:13227950

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 136)  
AUTHORS Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregorgis,E.,

Russell,D., de Jong,P. and Fraser,C.M.

Mouse BAC End Sequences from Library RPCI-24

Unpublished (1999)

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC

library availability, please contact Pieter de Jong

(pdejong@mail.cho.org). Clones may be purchased from BACPAC

Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end

page: http://www.tigr.org/tdb/bac.ends/mouse/bac\_end\_intro.html

Plate: 220 row: O column: 18

Seq primer: SP6

Class: BAC ends.

FEATURES

source

1. .136

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="RPCI-24-220018"

/clone\_lib="RPCI-24"

/sex="Male"

/cell\_type="Spleen/Brain"

/note="Vector: pFARBAC1; Site\_1: BamHI; Site\_2: BamHI;

RPCI-24 Mouse BAC Library produced by Pieter de Jong. The

library was cloned in the pFARBAC1 cloning vector at the

BamHI sites using MboI partially digested male C57BL/6J

DNA."

BASE COUNT 34 a 32 c 46 g 24 t

ORIGIN

alignment\_scores:

Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-701-711-2 x AZ909005

Align seg 1/1 to: AZ909005 from: 1 to: 136

769 LeuArgTyrSerAlaGlyValGly 776  
|||||  
50 CTCAGATACTCAGCGGTGTAGGG 73

seq\_name: gb\_est2:BF983453

seq\_documentation\_block:  
LOCUS BF983453 203 bp mRNA linear EST 23-JAN-2001  
DEFINITION 60230707F1 NIH\_MGC\_88 Homo sapiens CDNA clone IMAGE:4398427 5',  
mRNA sequence.  
ACCESSION BF983453  
VERSION BF983453.1 GI:12386265  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 203)  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM10100 row: i column: 20  
High quality sequence stop: 203.

FEATURES  
Location/Qualifiers  
1..203  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4398427"  
/clone\_lib="NIH\_MGC\_88"  
/tissue\_type="duodenal adenocarcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: small intestine; Vector: pCMV-SPORT6;  
Site 1: NotI; Site 2: SalI; Cloned unidirectionally;  
Oligo-dT primed. Average insert size 1.767 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: This is a NIH\_MGC Library."

BASE COUNT 20 a 65 c 94 g 24 t

ORIGIN

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-701-711-2 x BF983453 ..

Align seg 1/1 to: BF983453 from: 1 to: 203

667 SerSerLeuGlyProArgSerGln 674  
|||||  
95 AGCTCGTGCTGCCAAGGTCACAG 118

seq\_name: gb\_est1:AV257342

seq\_documentation\_block:  
LOCUS AV257342 206 bp mRNA linear EST 04-NOV-1999  
DEFINITION AV257342 RIKEN full-length enriched, adult male testis (DH10B) Mus musculus CDNA clone 492153N13 3', similar to AF053232 Mus musculus

BASE COUNT 52 a 53 c 32 g 69 t

ORIGIN

alignment\_scores:  
 Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-701-711-2 x AV257342/rev ..

Align seg 1/1 to reverse of: AV257342 from: 1 to: 206

227 ThrSerLeuGluAsnLeuArgAla 234

|||||

116 ACCTCCCTTGAAATTTGAGAGCA 93

seq\_name: gb\_est2:BM128835

seq\_documentation\_block:

LOCUS BM128835 208 bp mRNA linear EST 27-NOV-2001  
 DEFINITION If17c11.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens  
 cDNA 3' similar to SW:IDUA\_HUMAN P35475 ALPHA-L-IDURONIDASE  
 PRECURSOR ; mRNA sequence.

ACCESSION BM128835  
 VERSION BM128835.1 GI:17123387

KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 208)

REFERENCE  
 AUTHORS Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,  
 Lemishka, I., Scearce, M., Brestelli, J., Gradwohl, G., Clifton, S.,  
 Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A.,  
 Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas  
 M., Gibbons, M., McCann, R., Cole, R., Tsagarishvili, R., Williams, T.,  
 Jackson, Y. and Bowers, Y.

TITLE Endocrine Pancreas Consortium  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
 Endocrine Pancreas Consortium  
 Harvard University, Howard Hughes Medical Institute  
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
 MA 02138

Tel: 617-495-1812

Fax: 617-495-8557

Email: dmelton@biohp.harvard.edu

Library was constructed by Dr. Douglas Melton DNA sequencing by:  
 Washington University Genome Sequencing Center For information on  
 obtaining a clone please contact: Juliana Brown  
 (brownjfas.harvard.edu)

Trace considered overall poor quality

High quality sequence stop: 1.

Location/Qualifiers

1. .208

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="Melton Normalized Human Islet 4 N4-HIS 1"

/sex="Both"

/tissue\_type="Islets of Langerhans"

/dev\_stage="Adult"

/lab\_host="DH10B"

/note="Organ: Pancreas; Vector: pSPORT1; Site\_1: Not 1;  
 Site\_2: Sal 1; Starting library constructed using  
 SuperScript plasmid library kit (Life Technologies). cDNA  
 made by oligo-dT priming. Size-selected by column  
 fractionation; average insert size 1.08 kb. Library was  
 amplified once on solid support and plasmid DNA from  
 library was prepared. The library DNA was normalized by  
 method #4 from Bonaldo, Lennon, and Soares 1996 Genome  
 Research 6:791-806; 0.5 microgram single-stranded library

plasmid DNA was mixed with 5 micrograms PCR product  
 representing library inserts and hybridized to an EcoT of  
 20. Single-stranded (unhybridized) plasmids were isolated  
 by hydroxyapatite chromatography and used to make this  
 library."

BASE COUNT 38 a 39 c 91 g 38 t 2 others

ORIGIN

alignment\_scores:  
 Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-701-711-2 x BM128835/rev ..

Align seg 1/1 to reverse of: BM128835 from: 1 to: 208

591 LeuAspArgProValPheProThr 598

|||||

88 CTGGACCGACCGTCTTCCCCAG 65

seq\_name: gb\_est2:R70943

seq\_documentation\_block:

LOCUS R70943 215 bp mRNA linear EST 01-JUN-1995  
 DEFINITION Y150e05.s1 Soares placenta Nb2HP Homo sapiens cDNA clone  
 IMAGE:142688 3', mRNA sequence.

ACCESSION R70943

VERSION R70943.1 GI:844460

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 215)

REFERENCE  
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman  
 M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,  
 Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston  
 R., Williamson, A., Wohldmann, P. and Wilson, R.  
 The WashU-Merck EST Project  
 Contact: Willson RK  
 Unpublished (1995)

TITLE Washington University School of Medicine

JOURNAL 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

COMMENT Contact: Willson RK

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.wustl.edu

Insert Size: 1177

High quality sequence stops: 161

Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1177 Std Error: 0.00

Seq primer: Promega -21m13

High quality sequence stop: 161.

Location/Qualifiers

1. .215

/organism="Homo sapiens"

/db\_xref="GDB:551767"

/db\_xref="taxon:9606"

/clone="IMAGE:142688"

/clone\_lib="Soares placenta Nb2HP"

/sex="Female"

/dev\_stage="placenta obtained at birth (full term)"

/lab\_host="DH10B (ampicillin resistant)"

/note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a

modified polylinker; Site\_1: Not 1; Site\_2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer [5'

AAGTGAAGAAATTCGGCGCGAGGAATTTTTTTTTTTTTTTT 3'],

double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo. "

BASE COUNT 79 a 37 c 24 g 72 t 3 others

ORIGIN

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-701-711-2 x R70943 ..

Align seg 1/1 to: R70943 from: 1 to: 215

703 Leu1leuProLeuPropheLys 710

|||||

147 CTTATACCTCCACTCCCTTTAAA 170

seq\_name: gb\_est2:BF377641

seq\_documentation\_block:  
LOCUS BF377641 217 bp mRNA linear EST 24-NOV-2000  
DEFINITION PM1-TN0120-270800-001-al2 TN0120 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BF377641

VERSION BF377641.1 GI:11339666

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 217)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?ti=PM1st2=PM1-TN0120-270800-001-al2&t3=2000-08-27&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 2

High quality sequence stop: 217.

Location/Qualifiers

1..217

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="TN0120"

/dev\_stage="Adult"

/note="Organ: testis;normal; Vector: puc18; Site\_1: SmaI;

Site\_2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application

No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under

low stringency conditions."

BASE COUNT 74 a 32 c 40 g 71 t

ORIGIN

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-701-711-2 x BF377641 ..

Align seg 1/1 to: BF377641 from: 1 to: 217

250 LysLeuAsnIleAsnGluAspLys 257

|||||

162 AAGTTGAATATTAAATGAGGATAAA 185

seq\_name: gb\_est2:BI493603

seq\_documentation\_block:

LOCUS BI493603 233 bp mRNA linear EST 28-AUG-2001

DEFINITION df103c08.w1 Morton Fetal Cochlea Homo sapiens cDNA clone IMAGE:2540942 3', mRNA sequence.

ACCESSION BI493603

VERSION BI493603.1 GI:15332947

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 233)  
Robertson,N.G., Khetarpal,U., Gutierrez-Espeleta,G.A., Bieber,F.R. and Morton,C.C.

Isolation of novel and known genes from a human fetal cochlear cDNA library using subtractive hybridization and differential screening

Genomics 23, 42-50 (1994)

95130111

Contact: Morton, C. C.

Departments of Pathology and Obstetrics, Gynecology and

Reproductive Biology

Brigham and Women's Hospital

75 Francis Street, Harvard Medical School, Boston, MA 02115, USA

Tel: 617 732 7980

Fax: 617 738 6996

Email: cmorton@rics.bwh.harvard.edu

DNA sequencing and analyses were performed by National Institutes

of Health Intramural Sequencing Center (NISC; see

http://www.nisc.nih.gov).

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Plate: LLAM6329 row: F column: 15

Seq primer: T7 primer.

Location/Qualifiers

1..233

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="IMAGE:2540942"

/tissue\_type="Cochlea"

/dev\_stage="16-22 week fetus"

/lab\_host="SOLR cells (kanamycin resistant)"

/note="Organ: ear; Vector: pBluescript SK-; Site\_1: EcoRI;

Site\_2: XhoI; Reference: Genomics 23, 42-50 (1994) Cloned

unidirectionally. Primer: Oligo dT. Fetal cochlea, normal.

37% of inserts <0.5 kb, 56% 0.5-1.0 kb, 7% >1 kb. Uni-ZAP

XR Vector. Library constructed by N. Robertson, C. Morton.

-5' adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor

sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"

BASE COUNT 38 a 70 c 65 g 60 t

ORIGIN

alignment\_scores:



Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-701-711-2 x BI493603/rev ..

Align seg 1/1 to reverse of: BI493603 from: 1 to: 233

113 ProLysGluClyLeuCluGly 120

218 CCCAAGAAGGCTCCAGAAGGC 195

seq\_name: gb\_estl:AA729757

seq\_documentation\_block:

LOCUS AA729757 244 bp mRNA linear EST 27-JAN-1998  
 DEFINITION nx61d08.s1 NCI\_CGAP\_Al1v1 Homo sapiens cDNA clone IMAGE:1266735,  
 mRNA sequence.

ACCESSION AA729757

VERSION AA729757.1 GI:2751116

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 244)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: [cgaps-re@mail.nih.gov](mailto:cgaps-re@mail.nih.gov)

Tissue Procurement: Lee Helman, M.D., Michael R. Emmert-Buck, M.D.,  
 Ph.D.

cDNA Library Preparation: David B. Krizman, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone Distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Insert Length: 349 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham.

## FEATURES

source

1. .244

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:1266735"

/clone\_lib="NCI\_CGAP\_Al1v1"

/tissue\_type="alveolar rhabdomyosarcoma"

/lab\_host="DH10B"

/note="Vector: pAMP10; mRNA made from alveolar

rhabdomyosarcoma, cDNA made by oligo-dT priming.

Non-directionally cloned. Size-selected on agarose gel,

average insert size 600 bp. Reference: Krizman et al.

(1996) Cancer Research 56:5380-5383."

93 a 27 c 42 g 82 t

BASE COUNT

ORIGIN

## alignment\_scores:

Quality: 8.00 Length: 8

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-701-711-2 x AA729757/rev ..

Align seg 1/1 to reverse of: AA729757 from: 1 to: 244

703 LeuLeuProLeuProPhelys 710

|||||

36 CTTATACCTCCACTCCCTTTAAA 13

seq\_name: gb\_estl:BB417742

seq\_documentation\_block:

LOCUS BB417742

DEFINITION BB417742 RIKEN full-length enriched, 12 days embryo spinal cord Mus  
 musculus cDNA clone C530005A18 3', mRNA sequence.

ACCESSION BB417742

VERSION BB417742.1 GI:9239097

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

AUTHORS

1 (bases 1 to 253)

Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci  
 ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,  
 Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,  
 Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,  
 Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.,  
 Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,  
 , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata  
 ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugihara,Y.,  
 Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomimaga,N., Toya  
 ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.,  
 , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino  
 ,M., Muramatsu,M. and Hayashizaki,Y.  
 RIKEN Mouse ESTs (Konno,H., et al.)

## TITLE

JOURNAL

COMMENT

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: [genome-res@sc.riken.go.jp](mailto:genome-res@sc.riken.go.jp),

URL:<http://genome.gsc.riken.go.jp/>

Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki

,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Thermotabilization and thermoactivation of thermostable enzymes by

trehalose and its application for the synthesis of full length

cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)

Itoh,M., Kitsuunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,

Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki

,Y. and Hayashizaki,Y.

Automated filtration-based high-throughput plasmid preparation

system. Genome Res. 9 (5), 463-470 (1999)

Carninci,P. and Hayashizaki,Y.

High-efficiency full-length cDNA cloning. Methods Enzymol. 303,

19-44 (1999)

Please visit our web site (<http://genome.rtc.riken.go.jp>) for

further details.

Location/Qualifiers

1. .253

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone="C530005A18"

/clone\_lib="RIKEN full-length enriched, 12 days embryo

spinal cord"

/tissue\_type="spinal cord"

/dev\_stage="12 days embryo"

/lab\_host="DH10B"

/note="Site:1: Sali; Site:2: BamHI; cDNA library was

prepared and sequenced in Mouse Genome Encyclopedia

Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in

RIKEN. Division of Experimental Animal Research in Riken

contributed to prepare mouse tissues. 1st strand cDNA was

primed with a primer [5'

GAGAGAGAAGGATCCAGAGCTCTTTTTTTTTTTTTTTT 3'], cDNA was

prepared by using trehalose thermo-activated reverse

transcriptase and subsequently enriched for full-length by

cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGATTCGAGTTAAATTAATCCGCCGCCGCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 69 a 68 c 55 g 61 t

ORIGIN

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-701-711-2 x BB417742/rev ..

Align seg 1/1 to reverse of: BB417742 from: 1 to: 253

283 LeuThrTyrThrGlnAlaGluLeu 290  
|||||  
210 CTRACCTACACAGGCTGAATC 187

seq\_name: gb\_est1:BB579201

seq\_documentation\_block:  
LOCUS BB579201 258 bp mRNA linear EST 30-NOV-2000  
DEFINITION BB579201 RIKEN full-length enriched, 12 days embryo female mullerian duct Mus musculus cDNA clone 6820401C24 5', mRNA sequence.

ACCESSION BB579201  
VERSION BB579201.1 GI:11475148  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 258)

TITLE Aizawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T., Carninci,P., Hanagaki,T., Hayatsu,N., Hiraoka,T., Hirozane,T., HodoYama,Y., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kawai,J., Kojima,Y., Konno H., Kusakabe,M., Matsuyama,T., Miyazaki,A., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Okazaki,Y., Okido,T., Owa,C., Sakai,C., Sakai,K., Sasaki,D., Sato,K., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Tanaka T., Toya,T., Watahiki,A., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshiki,A., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Aizawa,K. et al. 2000)  
Unpublished (2000)

JOURNAL Contact: Yoshihide Hayashizaki  
COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp,  
URL:http://genome.gsc.riken.go.jp/  
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Thermotabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki Y. and Hayashizaki,Y.  
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)  
Carninci,P. and Hayashizaki,Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)  
Please visit our web site (<http://genome.ritc.riken.go.jp>) for further details.

FEATURES  
source

Location/Qualifiers  
1. .258  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="6820401C24"  
/clone\_lib="RIKEN full-length enriched, 12 days embryo female mullerian duct"  
/sex="female"  
/tissue\_type="mullerian duct includes surrounding region"  
/dev\_stage="12 days embryo"  
/lab\_host="DH10B"  
/note="Site 1: Sali; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGATTCGAGTTAAATTAATCCGCCGCCGCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: Sali; 3' end: BamHI"

BASE COUNT 93 a 42 c 77 g 46 t

ORIGIN

## alignment\_scores:

Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-701-711-2 x BB579201 ..  
Align seg 1/1 to: BB579201 from: 1 to: 258

230 GluAsnLeuArgAlaLysTyrLeu 237  
|||||

155 GAAACCTTACGGCAAGTACTTA 178

seq\_name: gb\_est1:AA247205

## seq\_documentation\_block:

LOCUS AA247205 262 bp mRNA linear EST 11-MAR-1997  
DEFINITION csq0312.seq.F Human fetal heart, Lambda ZAP Express Homo sapiens cDNA 5', mRNA sequence.

ACCESSION AA247205

VERSION AA247205.1 GI:1878590

KEYWORDS EST.

SOURCE human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 262)

AUTHORS Liew,C.C.

TITLE cDNAs from human fetal heart (1997)

JOURNAL Unpublished (1997)

COMMENT Contact: Liew CC

Brigham and Women's Hospital

Harvard Medical School

75 Francis St. Boston, MA 02115, USA

Tel: 6177328915

Fax: 6179750995

Email: [cliew@rics.bwh.harvard.edu](mailto:cliew@rics.bwh.harvard.edu)

PCR Primers

FORWARD: 5' GCCAGCTCGAATTAACCCCTCACTAAAGG 3'

BACKWARD: 5' CCAGTGAATTGATACGACTCACTATAGGCG 3'

Seq primer: 5' GAAATTAACCTCCTACTAAAGG 3'.

```

FEATURES
  source
    1. .262
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone_lib="Human fetal heart, Lambda ZAP Express"
      /lab_host="E. coli XL1-Blue"
      /note="Vector: Lambda ZAP Express; Site_1: EcoRI; Site_2:
XhoI; mRNA was purified from human fetal hearts (8-10
weeks). cDNA was synthesized using a XhoI-Oligo dT
adaptor-primer. EcoRI adaptors were ligated, followed by
digestion with XhoI, for directional cloning into
predigested lambda ZAP Express."
    70 a 58 c 63 g 71 t
    BASE COUNT
    ORIGIN

alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
  US-09-701-711-2 x AA247205 ..
  Align seg 1/1 to: AA247205 from: 1 to: 262

693 GlyAsnAlaLeuAlaThrPheGly 700
|||||
233 GCCAATGCATTGCCAACCTTTGGT 256

seq_name: gb_est1:AI866213

seq_documentation_block:
  LOCUS AI866213 268 bp mRNA linear EST 07-MAR-2000
  DEFINITION W127c10.x1 NCI_CGAP_Utl1 Homo sapiens cDNA clone IMAGE:2426130 3',
  mRNA sequence.
  ACCESSION AI866213 GI:5530320
  VERSION AI866213.1
  KEYWORDS EST.
  SOURCE human.
  ORGANISM Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE 1 (bases 1 to 268)
  AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index
  JOURNAL Unpublished (1997)
  COMMENT Contact: Robert Strausberg, Ph.D.
  Email: cgapbs-r@mail.nih.gov
  Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
  Emmert-Buck, M.D., Ph.D.
  cDNA Library Preparation: Life Technologies, Inc.
  cDNA Library Arrayed by: Greg Lennon, Ph.D.
  DNA Sequencing by: Washington University Genome Sequencing Center
  Clone distribution: NCI-CGAP clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  www-bio.llnl.gov/bbrp/image/image.html
  Insert length: 855 Std Error: 0.00
  Seq primer: -400P from Gibco
  High quality sequence stop: 211.
  Location/Qualifiers
    1. .268
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone_lib="NCI_CGAP_Utl1"
      /tissue_type="well-differentiated endometrial
adenocarcinoma, 7 pooled tumors"
      /lab_host="DH10B"
      /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.

```

Average insert size 1.75 kb. Life Technologies catalog #:  
11538-014"

BASE COUNT 98 a 46 c 27 g 97 t  
ORIGIN

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-701-711-2 x AI866213 ..

Align seg 1/1 to: AI866213 from: 1 to: 268

703 LeuIleLeuProLeuProPhelys 710  
|||||  
198 CTTATACTTCCACCTCCCTTAA 221

seq\_name: gb\_est1:AI583432

seq\_documentation\_block:

LOCUS AI583432 274 bp mRNA linear EST 14-DEC-1999  
DEFINITION tr-96h06.x1 NCI\_CGAP\_Pan1 Homo sapiens cDNA clone IMAGE:2226971 3',  
mRNA sequence.

ACCESSION AI583432  
VERSION AI583432.1 GI:4569329  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 274)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Life Technologies catalog #: 11548-013

DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Insert length: 2636 Std Error: 0.00

Seq primer: -400P from Gibco

High quality sequence stop: 197

POLYA-No.

Location/Qualifiers

FEATURES

source

1. .274

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="IMAGE:2226971"

/clone\_lib="NCI\_CGAP\_Pan1"

/tissue\_type="adenocarcinoma"

/lab\_host="DH10B"

/note="Organ: pancreas; Vector: pCMV-SPORT6; Site\_1: SalI;

Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.72 kb. Life Technologies catalog #:

11548-013"

BASE COUNT 97 a 43 c 27 g 107 t

ORIGIN

alignment\_scores:

Quality: 8.00 Length: 8

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-701-711-2 x AI583432 ..



## Laboratory Mouse DNA Resource

(<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g114732114[gblAF129072.1]), a copy-number inducible derivative of plasmid RL. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

BASE COUNT 67 a 60 c 55 g 95 t  
ORIGIN

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-701-711-2 x AZ375019 ..  
Align seg 1/1 to: AZ375019 from: 1 to: 277

703 LeuileLeuProLeuProPhelys 710  
|||||  
124 CTAATCTTGCCTTTCCTTCAG 147

seq\_name: gb\_est2:H82436

seq\_documentation\_block:  
LOCUS H82436 293 bp mRNA linear EST 09-NOV-1995  
DEFINITION YU80q04.s1 Soares fetal liver spleen lNFLS Homo sapiens cDNA clone IMAGE:240150 3', mRNA sequence.

ACCESSION H82436  
VERSION H82436.1 GI:1060525  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 293)  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.  
The WashU-Merck EST Project

TITLE The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.wustl.edu

Insert Size: 1189  
High quality sequence stops: 281  
Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Insert Length: 1189 Std Error: 0.00  
Seq primer: Promega -21ml3  
High quality sequence stop: 281.  
Location/Qualifiers

FEATURES  
source  
1..293  
/organism="Homo sapiens"  
/db\_xref="GDB:3789103"

/db\_xref="taxon:9606"  
/clone="IMAGE:240150"  
/clone\_lib="Soares fetal liver spleen lNFLS"  
/sex="male"  
/dev\_stage="20 week-post conception fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia) with a modified polylinker; Site: 1: Pac 1; Site 2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGGAGAAATTAATTAAGATCTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 93 a 51 c 41 g 102 t  
ORIGIN

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-701-711-2 x H82436 ..  
Align seg 1/1 to: H82436 from: 1 to: 293

703 LeuileLeuProLeuProPhelys 710  
|||||  
186 CTTATCTTCCACTTCCCTTTAAG 209

seq\_name: gb\_est1:AV205177

seq\_documentation\_block:  
LOCUS AV205177 295 bp mRNA linear EST 30-OCT-1999  
DEFINITION AV205177 RIKEN full-length enriched, adult male testis Mus musculus cDNA clone 1700073108 3', mRNA sequence.

ACCESSION AV205177  
VERSION AV205177.1 GI:6146030  
KEYWORDS EST.  
SOURCE house mouse.

ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
REFERENCE 1 (bases 1 to 295)  
Konno, H., Alzawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Takahashi, F., Tateno, M., Tominaga, N., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

RIKEN Mouse ESTs (Konno, H., et al. 1999)  
Unpublished (1999)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: [genome-res@sc.riken.go.jp](mailto:genome-res@sc.riken.go.jp),  
URL: <http://genome.gsc.riken.go.jp/>  
Sasaki, N., Izawa, M., Watahiki, M., Ozawa, K., Tanaka, T., Yoneda, Y., Matsuura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

TITLE RIKEN Mouse ESTs (Konno, H., et al. 1999)  
JOURNAL Unpublished (1999)  
COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: [genome-res@sc.riken.go.jp](mailto:genome-res@sc.riken.go.jp),  
URL: <http://genome.gsc.riken.go.jp/>  
Sasaki, N., Izawa, M., Watahiki, M., Ozawa, K., Tanaka, T., Yoneda, Y., Matsuura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)



```
/dev_stage="11 weeks old"
/lab_host="SOLR (kanamycin resistant)"
/notes="Organ: skin; Vector: pBluescript SK-; Site_1: EcoRI
; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo
dr. Whole skin from 11 week old C57BL/6 female mice.
Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"
```

```
BASE COUNT      64 a      70 c      60 g      108 t
ORIGIN
```

```
alignment_scores:
  Quality:      8.00      Length:      8
  Ratio:      1.000      Gaps:      0
Percent Similarity: 100.000      Percent Identity: 100.000
```

```
alignment_block:
US-09-701-711-2 x AI528822/rev ..
Align seg 1/1 to reverse of: AI528822 from: 1 to: 302
```

```
705 LeuProLeuProPheLysGlyAsp 712
|||||
90 CTGCCCCCTACCCCTCAGGAGAT 67
```







1 MetArgAsnSerTyrPheLysGlyPheGlnValSerAlaMetThrMetAl 17  
|||||  
1 ATCGCTAATTCATATTTAAAGGTTTTCAGGTCAGTCAATGACAAATGGC 50  
17 avaI MetMetValMetSerThrHisAlaGlnAlaAlaAspPheMetAlaa 34  
|||||  
51 TGTCTATGATGGTAAATGTCAACCTCATGCAACAAGCGCGGATTTTATGGCAA 100  
34 snAspIleThrIleThrGlyLeuGlnArgValThrIleGluSerLeuGln 50  
|||||  
101 ATGACATTTACCATCACAGACTACAGCGAGTGACCATTTGAAGCTTTACAA 150  
51 SerValLeuProPheArgLeuGlyGlnValValSerGluAsnGlnLeuAl 67  
|||||  
151 AGCGTGCTGCCGTTTCGCTTGGGTCAAGTGTGAGCGAAACACGATTGGC 200  
67 aAspGlyValLysAlaLeuTyrAlaThrGlyAsnPheSerAspValGlnV 84  
|||||  
201 TGTATGGTGTCAAGACACTTTATGCAACAGGCAATTTTTCAGATGTGCAAG 250  
84 alTyrHisGlnGluGlyArgIleIleTyrGlnValThrGluArgProLeu 100  
|||||  
251 TCTATCATCAAGAGGGCGTATCATCTATCAGGTAACCGAAGGCCGTTA 300  
101 IleAlaGluIleAsnPheGluGlyAsnArgLeuIleProLysGluGlyLe 117  
|||||  
301 ATCGCTGAGATTAATTTTGGGGCAATCGCTTAATTCCAAAGAAAGGTC 350  
117 uGlnGluGlyLeuLysAsnAlaGlyLeuAlaValGlyGlnProLeuLysG 134  
|||||  
351 ACAAGAAGGGCTAAATAATGCTGGCTTACGTGGGTCAACCACTAAAC 400  
134 InAlaThrValGlnMetIleGlnThrGluLeuThrAsnGlnTyrIleSer 150  
|||||  
401 AAGCCACAGTACAGATGATCAACACCGAGCTTACCAATCAATATATATCA 450  
151 GlnGlyTyrTyrAsnThrGluIleThrValLysGlnThrMetLeuAspGl 167  
|||||  
451 CAAGGCTATTATATACCGAAATTAATCTCAACAGACGATGCTTGATGG 500  
167 yAsnArgValLysLeuAspMetThrPheAlaGluGlyLysProAlaArgV 184  
|||||  
501 TAATCGTGTAAAGCTTGATATGACCTTTGCTGAAGGTAACCTGCACGGG 550  
184 alValAspIleAsnIleIleGlyAsnGlnHisPheSerAspAlaAspLeu 200  
|||||  
551 TGGTGTATTAATATATCATTTGGCAATCAGCATTTTAGCGATGCAGATTG 600  
201 IleAspValLeuAlaIleLysAspAsnLysIleAsnProLeuSerLysAl 217  
|||||  
601 ATTCATGTGCTTGGCATTAAGGATAATAAATCAATCCACCTGTCTAAAGC 650  
217 aAspArgTyrThrGlnGluLysLeuValThrSerLeuGluAsnLeuArgA 234  
|||||  
651 TGACGGTTATATCTCAAGAAAAGCTGGTGACCATTTAGAGAAATTTGGCTG 700  
234 laLysTyrLeuAsnAlaGlyPheValArgPheGluIleLysAspAlaLys 250  
|||||  
701 CTAATATCTCAATGCAGGGTTTGGCGCTTTTGAGATTAAAGATGCTAAG 750  
251 LeuAsnIleAsnGluAspLysAsnArgIlePheValGluIleSerLeuHi 267  
|||||  
751 CTTAATATTATTAAGATAAAACCGTATCTTGTGTGAGATTTTCATTGCA 800  
267 sGluGlyGlnTyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuT 284  
|||||  
801 TGAAGGTGAGCAATATCGCTTTGGACAGACACAGCTTTTGGGTAAATTTAA 850  
284 hrTyrThrGlnAlaGlnLeuGluAlaLeuLeuLysPheLysAlaGluGlu 300  
|||||  
851 CTTATACTCAAGCAAGAACTTCAGGCACTGCTTAATTCAAAGCAGAGAA 900  
301 GlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLy 317  
|||||

901 GGGTTTTCAACAGCCATGCTTGAGCAAAACAACAATATCAGTACCAA 950  
317 sPheGlyAspAspGlyTyrTyrTyrAlaGlnIleArgProValThrArgI 334  
|||||  
951 ATTTGGTGACGATGGCTATTATTATGCTCAAAATCCGTCCTTAACACGCA 1000  
334 leAsnAspGluSerArgThrValAspValGluTyrTyrIleAspProVal 350  
|||||  
1001 TTAATGATGAAGTCGTACGGTTGATGTGGAATATTATTGACCCGTGA 1050  
351 HisProValTyrValArgIleAsnPheThrGlyAsnPheLysThrGl 367  
|||||  
1051 CACCCTGTCTATGTAGCCGCTATTAAATTTTACAGTAACCTTTAAAGACCA 1100  
367 nAspGluValLeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlas 384  
|||||  
1101 AGATGAAGTACTCGCTCGTGAATGCGCAACCTTGAAGGTGCGTTGGCAT 1150  
384 erAsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhe 400  
|||||  
1151 CTAATCAAAAATCCAGCTGCTCGTCGACGCTTGATGCGGACTGGGTTT 1200  
401 PheLysHisValThrValAspThrArgProValProAsnSerProAspGl 417  
|||||  
1201 TTTAAACATGTTACCGTTGATCTCGTCAGTACCCCACTCACCTGATCA 1250  
417 nValAspValAsnPheValValGluGlnProSerGlySerSerThrI 434  
|||||  
1251 GGTGTGATGTAATTTTGGTTGAAGAACAACTTCAGGATCATCAACCA 1300  
434 leAlaAlaGlyTyrSerGlnSerGlyValThrPheGlnPheAspVal 450  
|||||  
1301 TCGCAGCAGGCTACTCTCAAAGTGGTGGTAACTTTTCAATTTGATGTT 1350  
451 SerGlnAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSe 467  
|||||  
1351 TCTCAAAATAACTTTATGGGTACAGGTAAAGCACGCTCAATGCTTGGTTTC 1400  
467 rArgSerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrP 484  
|||||  
1401 TCGCTCTGAGACCCGCTGAGGTGTATAGTTTGGGTATGACCAACCCATACT 1450  
484 heThrValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThr 500  
|||||  
1451 TTACCGTAAATGGCGTCTCGCAAGCTTGAGTGGCTACTATCGCAAAACC 1500  
501 LysTyrAspAsnLysAsnIleSerAsnTyrValLeuAspSerTyrGlyGl 517  
|||||  
1501 AAGTATGATAACAAGAACATTAGTAATTTATGTACTTGTATTCTTATGGTGG 1550  
517 ySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSerPheG 534  
|||||  
1551 CTCATTAAAGCTATGGATATCCAAATGTAGTAAATCAACGCAATAAGCTTGG 1600  
534 lyLeuAsnAlaAspAsnThrLysLeuHisGlyGlyArgPheMetGlyIle 550  
|||||  
1601 GTCTGAATGCTGACAATACCAAGCTTCATGCGCGTCTTTTATGGGCATT 1650  
551 SerAsnValLysGlnLeuMetAlaAspGlyLysIleGlnValAspAs 567  
|||||  
1651 AGTAATGTCAAGCAGCTGATGGCAGATGGTGGCAAAATTCAGTGGATAA 1700  
567 nAsnGlyIleProAspPheLysHisAspTyrThrThrAsnAlaIleL 584  
|||||  
1701 TAAATGGCATTCCTGATTTTAAAGCATGATTACAAACCTTACAAATGCCATT 1750  
584 euGlyTyrAsnTyrSerSerLeuAspArgProValPheProThrGlnGly 600  
|||||  
1751 TGGGGTGAATTTTCAAGTCTAGATCGCCCTGTATTTCCAACCAAGGC 1800  
601 MetSerHisSerValAspLeuThrValGlyPheGlyAspLysThrHisGl 617  
|||||

1801 ATGAGTCATTCTAGATTGACGGTTGGTTTGGTCATAAAAACTCATCA 1850  
 617 nLysValValTyrGlnGlyAsnIleTyrArgProPheIleLysSerV 634  
 |||||  
 1851 AAAAGTGGTTTATCAAGGAATATCATCGCCATTATCAAAAAATCAG 1900  
 634 aLeuArgGlyTyrAlaLysLeuGlyTyrGlyAsnLeuProPheTyr 650  
 |||||  
 1901 TCTTGGTGGATACGCCAAGTAGGTAGGCAATAATTACCATTTAT 1950  
 651 GluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667  
 |||||  
 1951 GAAATTTCTATGACGGCGCTATGTTGGTTCGGTTCGTGCTATGATCAATC 2000  
 667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgGlyGlnG 684  
 |||||  
 2001 CTCCTTGGGTCCAGCTCAAGCCATTATTGACAGCTCGTGGTGGTCAAC 2050  
 684 lnThrThrLeuGlyGluValValGlyGlyAsnAlaLeuAlaThrPheGly 700  
 |||||  
 2051 AAACACACTAGGAGAGGTGTTGGTGGTAATGCTTTGGCAACTTTCGGC 2100  
 701 SerGluLeuLeuProLeuProPheLysGlyAspTrpIleAspGlnVa 717  
 |||||  
 2101 AGTGAGCTGATTTTACCTTGGCCATTAAAGGTGATTTGGATAGATCAGGT 2150  
 717 lArgProValIlePheIleGluGlyGlnValPheAspThrThrGlyM 734  
 |||||  
 2151 GCGTCAGTGATATTCATTGAGGGCGGTGAGTTTGTATACACAGGTA 2200  
 734 eAspLysGlnThrIleAspLeuThrGlnPheLysAspProGlnAlaThr 750  
 |||||  
 2201 TGGATAAACAAACCATTTGATTTAAACCAATTTAAAGACCACCAAGCAACA 2250  
 751 AlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuThrGlnAspLy 767  
 2251 GTTGAACAAATGCAAAAGCAGCCAAATCCCGCTACTAACCCCAAGATAA 2300  
 767 sGlnLeuArgTyrSerAlaGlyValGlyAlaThrTrpTyrThrProIleG 784  
 |||||  
 2301 ACAGTTGCGTTATAGTGTGTTGGTGCACACTTGTATACGCCCATTTG 2350  
 784 lProLeuSerIleSerTyrAlaLysProLeuAsnLysLysGlnAsnAsp 800  
 |||||  
 2351 GTCCCTTATCTATTAGCTATGCAAGCCATTGAATAAAAAACAATAATGAT 2400  
 801 GlnThrAspThrValGlnPheGlnIleGlySerValPhe 813  
 |||||  
 2401 CAGACCGATACGGTACAGTTCAGATTTGGTGTGCTCTTT 2439

seq\_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAZ29551

seq\_documentation\_block:

ID: AAZ29551 standard; DNA; 2442 BP.

XX AC AAZ29551;

XX AC AAZ29551;

DT 14-MAR-2000 (first entry)

XX DE M. catarrhalis (ATCC 43617) BASB027 polypeptide encoding DNA #2.

XX KW BASB027; OMP85; outer membrane protein; otitis media; diagnosis; treatment; screening; bacterial infection; ds.

XX OS Moraxella catarrhalis.

XX FH Key Location/Qualifiers

FT CDS 1..2442

FT /\*tag= a

FT /product= "BASB027 polypeptide"

XX PN W09963093-A2.

XX

PD 09-DEC-1999.  
 XX 31-MAY-1999; 99WO-EP03822.  
 XX 03-JUN-1998; 98GB-0011945.  
 PR 08-MAR-1999; 99GB-0005304.  
 XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 XX Vinals-Bassols C;  
 XX WPI; 2000-105700/09.  
 DR P-PSDB; AAY44391.  
 XX Novel BASB027 polynucleotide and polypeptides from Moraxella  
 PT catarrhalis useful for treating M. catarrhalis infection such as otitis  
 PT media  
 XX Claim 8; Page 102; 109pp; English.  
 XX The present sequence is a DNA obtained from chromosomal DNA library of  
 CC Moraxella catarrhalis strain Mc2931 (ATCC 43617). It encodes BASB027  
 CC polypeptide, which shows significant homology to Neisseria meningitidis  
 CC OMP85 outer membrane protein. BASB027 polynucleotide and polypeptide can  
 CC be used for diagnosis and staging of disease, determining susceptibility  
 CC to a disease and to prepare medicaments for treating M. catarrhalis  
 CC infections, especially otitis media. The BASB027 DNA can be used as  
 CC probe for screening of genetic mutations, serotype, taxonomic  
 CC classification or identification. BASB027 agonists, antagonists and  
 CC antibodies may be used to prevent and/or treat bacterial infections.  
 XX SQ Sequence 2442 BP; 735 A; 461 C; 535 G; 711 T; 0 other;

alignment\_scores:

Quality: 4189.00 Length: 813

Ratio: 5.159 Gaps: 0

Percent Similarity: 99.877 Percent Identity: 99.754

alignment\_block:

US-09-701-711-2 x AAZ29551 ..

Align seg 1/1 to: AAZ29551 from: 1 to: 2442

1 MetArgAsnSerTyrPheLysGlyPheGlnValSerAlaMetThrMetAl 17  
 |||||  
 1 ATGCGTAATTATATTTTAAAGGTTTTCAGTGCATGCAATGACATGGC 50  
 17 aValMetMetValMetSerThrHisAlaGlnAlaAlaAspPheMetAlaA 34  
 |||||  
 51 TGTCATGATGGTAAATGCTCAACTCATGCACAAAGCGCGGATTTTATGCAA 100  
 34 snAspIleThrIleThrGlyLeuGlnArgValThrIleGluSerLeuGln 50  
 |||||  
 101 ATGACATTGCCATCACAGGACTACAGGAGTGACCATTTGAAAGCTTACAA 150  
 51 SerValLeuProPheArgLeuGlyGlnValValSerGluAsnGlnLeuAl 67  
 |||||  
 151 ACGGTGCTCCGCTTCGCTTGGTCAAGTGGTGGAGGCAAGCACAGTTGGC 200  
 67 aAspGlyValLysAlaLeuTyrAlaThrGlyAsnPheSerAspValGlnV 84  
 |||||  
 201 TGATGCTGTCAAAGCACTTTATGCAACAGGCAATTTTTCAGATGTGCAAG 250  
 84 alTyrHisGlnGluGlyArgIleIleTyrGlnValThrGluArgProLeu 100  
 |||||  
 251 TCTATCATCAAGAAAGGCGGTATCATCTATCAGGTAACCCGAAAGGCGGTTA 300  
 101 IleAlaGluIleAsnPheGluGlyAsnArgLeuIleProLysGluGlyLe 117  
 |||||  
 301 ATCGCTGAGATTAATTTTTCAGGCAATCGCTTAATTCACAAAGAGGTCT 350  
 117 uGlnGluGlyLeuLysAsnAlaGlyLeuAlaValGlyGlnProLeuLysG 134

```

|||||
351 ACAGAGGGCTAAAAATGCTGGCTAGCTGTGGTCAACCACTAAAAAC 400
134 lnAlaThrValGlnMetIleGluThrGluLeuThrAsnGlnTyrIleSer 150
|||||
401 AAGCCACAGTACAGATGATCGAAACCGAGCTTACCAATCAATATATCA 450
151 GlnGlyTyrTyrAsnThrGluIleThrValLysGlnThrMetLeuAspG1 167
|||||
451 CAAGGCTATTATATACCGAAATTACTGTCAACACAGAGATGCTTGATGG 500
167 yAsnArgValLysLeuAspMetThrPheAlaGluGlyLysProAlaArgV 184
501 TAATCGTGTAAAGCTTGATATGACCTTTGCTGAAGGTAAACCTGCACGG 550
184 alValAspIleAsnIleIleGlyAsnGlnHisPheSerAspAlaAspLeu 200
551 TGGTGTGATATTAAATATCATTTGGCAATCAGCATTTTAGCGATGCGATTG 600
201 IleAspValLeuAlaIleLysAspAsnLysIleAsnProLeuSerLysAl 217
601 ATTGATGTCTTGGATTAAGGATAATAAATCAATCCAC1GTCTAAAGC 650
217 aAspArgTyrThrGlnGluLysLeuValThrSerLeuGluAsnLeuArgA 234
651 TCACCGTTTACTCAAGAAAAGCTGGTGACCAGTTTACAGAAATTTGCGTG 700
234 laLysTyrLeuAsnAlaGlyPheValArgPheGluIleLysAspAlaLys 250
701 CTAAATATCTCAATGCAGGGTTTGTGCGCTTTTGAGATTAAAGATGCTAAG 750
251 LeuAsnIleAsnGluAspLysAsnArgIlePheValGluIleSerLeuHi 267
751 CTTAATATTATGAGAGATAAAACCGTATCTTTGTTGAGATTTCAATGCA 800
267 sGluGlyGluGlnTyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuT 284
801 TGAAGGTGAGCAATATCGCTTGGACAGACACAGTTTTTTGGGTAATTTAA 850
284 hrTyrThrGlnAlaGluLeuGluAlaLeuLeuLysPheLysAlaGluGlu 300
851 CTTATACTCAAGCAACTTGAGGCAC1TGCCTTAAATTCAAAGCAGAAGAA 900
301 GlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLy 317
901 GGGTTTTCAACAGCCATGCTTGAGCAAAACAAACAATATCAGTACCAA 950
317 sPheGlyAspAspGlyTyrTyrTyrAlaGlnIleArgProValThrArgI 334
951 ATTTGGTGACGATGGCTATTATTATGCTCAAAATCCGCTCTGTAACACGCA 1000
334 leAsnAspGluSerArgThrValAspValGluTyrTyrIleAspProVal 350
1001 TTAATGATGAAGACTCGTACGGTTGATGGATATATTATGACCCCTGTA 1050
351 HisProValTyrValArgIleAsnPheThrGlyAsnPheLysThrG1 367
1051 CACCCTGTCTATGTACGGCGTATTAAATTTACAGGTAACTTTAAGACCCA 1100
367 nAspGluValLeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaS 384
1101 AGATGAAGTACTCCGTCGTGAGATCGCAACCTTGAAGGTGCGTTGGCAT 1150
384 erAsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhe 400
1151 CTAATCAAAAAATCCAGCTGCTCGTGACGCTTGATGGGACTGGGTTT 1200
401 PheLysHisValThrValAspThrArgProValProAsnSerProAspG1 417
1201 TTTAAACATGTTACCGTTGATCTCGTCAGTACCACCACTCACCTCATCA 1250
417 nValAspValAsnPheValValGluGlnProSerGlySerSerThrI 434
|||||

```

```

1251 GGTGTGATCTAAATTTTGTGGTTGAAGAACCAACCTTCAGGATCATCAACCA 1300
434 leAlaAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspVal 450
1301 TCGCAGCAGGCTACTCTCAAAAGTGGTGGTGAACCTTTTCAATTTGATGTT 1350
451 SerGlnAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSe 467
1351 TCTCAAAATAACTTTATGGGTACAGGTAAAGCACGTCGAATGCTTCGTTTC 1400
467 rArgSerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrP 484
1401 TCGCTCTGAGACCGCTGAGGTGTATAGTTTGGGTATGACCAACCCATAT 1450
484 heThrValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThr 500
1451 TTACCGTAAATGGGCTCTCGCAAGCTTGAGTGGCTACTATCGTAAACCC 1500
501 LysTyrAspAsnLysAsnIleSerAsnTyrValLeuAspSerTyrGlyG1 517
1501 AAGTATGATAACAAGAACATTAGTAATTATCTACTTGTATTCTTATGGTGG 1550
517 ySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSerPheG 534
1551 CTCATTAAAGCTATGGATATCAATTTGATGAAAATCAACGCATAAGCTTTG 1600
534 lyLeuAsnAlaAspAsnThrLysLeuHisGlyGlyArgPheMetGlyIle 550
1601 GTCGAATGCTGACAAATACCAGCTTCATGGCGGTCTGTTTATGGGCATT 1650
551 SerAsnValLysGlnLeuMetAlaAspGlyLysIleGlnValAspAs 567
1651 AGTAATGTCAAGCAGCTGATGCGCAGATGGTGCAAAATTCGAATGGATAA 1700
567 nasnGlyIleProAspPheLysHisAspTyrThrThrTyrAsnAlaIleL 584
1701 TAATGGCATTCCTGATTTTAAAGCATGATTACACAACCTACAAATGCCATT 1750
584 euGlyTyrAsnTyrSerSerLeuAspArgProValPheProThrGlnGly 600
1751 TGGGTGCAATTAITCAAGTCTAGATCGCCTGTATTTCCAACCCCAAGGC 1800
601 MetSerHisSerValAspLeuThrValGlyPheGlyAspLysThrHisG1 617
1801 ATGAGTCATCTCTAGATTTGACGGTGTGGTTTGGTGATAAAACCTCATCA 1850
617 nLysValValTyrGlnGlyAsnIleTyrArgProPheIleLysLysSerV 634
1851 AAAAGTGGTTTATCAAGGCAATATCTATCGCCCATTTATCAAAAATCAG 1900
634 alLeuArgGlyTyrAlaLysLeuGlyTyrGlyAsnAsnLeuProPheTyr 650
1901 TCTTGGTGGATACGCCAAGTTAGGCTATGGCAATAATTTACCATTTTAT 1950
651 GluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667
1951 GAAAAATTTCTATGACGGCGGTATGGTTCGGTTCGGTCTATGATCAATC 2000
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
2001 CTCCTTTGGGTCCACGCTCACAAGCCTATTGACAGCTCGTCTGTGTCAC 2050
684 lnThrThrLeuGlyGluValValGlyGlyAsnAlaLeuAlaThrPheGly 700
2051 AAACCCACACTAGGAGAGGTTGTTGGTGGTAAATGCTTTGGCAACTTTCCGC 2100
701 SerGluLeuIleLeuProIleProPheLysGlyAspTyrIleAspGlnVa 717
2101 AGTGAGCTGATTTTACCTTTGCCATTTAAAGGTGATTGGATAGATCAGGT 2150
717 lArgProValIlePheIleGluGlyGlyGlnValPheAspThrThrGlyM 734
2151 GCGTCCAGTGATATTCAATTGAGGCGGTGAGGTTTGTGATACAACAGGTA 2200

```

734 etAspLysGlnThrIleAspLeuThrGlnPheLysAspProGlnAlaThr 750  
|||||  
2201 TGGATTAACAACACCATTTGATTTAAACCAATTTAAAGACCCACACGACACA 2250  
|||||  
751 AlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuThrGlnAspLys 767  
|||||  
2251 GCTGAACAAAATGCAAAAGCAGCCCAATGCCCGCTACTAACCCCAAGATAA 2300  
|||||  
767 scLLeuArgTyrSerAlaGlyValGlyAlaThrTrpTyrThrProIleG 784  
|||||  
2301 ACAGTGTGGTATATAGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2350  
|||||  
784 lyProLeuSerIleSerTyrAlaLysProLeuAsnLysLysGlnAsnAsp 800  
|||||  
2351 GTCCCTTATCTATTAGCTATGCCAAGCCATTGAATAAAAACAATAATGAT 2400  
|||||  
801 GlnThrAspThrValGlnPheGlnIleGlySerValPhe 813  
|||||  
2401 CAGACCGATACGGTACAGTTCAGATTCCAGATTGGTAGTGCTTTT 2439  
|||||

seq\_name: /SIDS1/gcgdata/geneseq/geneseq-embl/NA2001A.DAT:AAF28550

seq\_documentation\_block:

ID AAF28550 standard; DNA; 99629 BP.

XX AC AAF28550;

XX DT 04-APR-2001 (first entry)

XX XX Genomic fragment #37.

XX DE Genomic library; bacteria; human upper airway; otitis media; sinusitis;  
XX KW bronchopulmonary; endocarditis; meningitis; ss.

XX OS Moraxella catarrhalis.

XX PN WO200078968-A2.

XX PD 28-DEC-2000.

XX PF 16-JUN-2000; 2000WO-US16649.

XX PR 18-JUN-1999; 99US-0140121.

XX PA (INCY-) INCYTE GENOMICS INC.

XX PI Lagace RE, Patterson C, Berg KL;

XX WP1; 2001-041427/05.

XX XX Genomic library for identifying diagnostic and therapeutic  
XX PT compositions, and for identifying virulence factors, regulatory  
XX PT elements and drug targets, comprises Moraxella catarrhalis nucleic  
XX PT acids -

XX PS Claim 1; Page 391-415; 545pp; English.

XX CC The present invention relates to a Moraxella catarrhalis genomic library  
XX CC comprising of a combination of 41 nucleic acid molecules (see  
XX CC AAF28514-AAF28554). The library has a number of uses described in the  
XX CC specification e.g. is useful for identifying diagnostic and therapeutic  
XX CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large  
XX CC aerobic, gram-negative diplococcus, normally found among the bacterial  
XX CC flora of human upper airways. M. catarrhalis is known to cause acute,  
XX CC localised infections such as otitis media, sinusitis and bronchopulmonary  
XX CC infection and life-threatening, systemic diseases including endocarditis  
XX CC and meningitis.

XX SQ Sequence 99629 BP; 29233 A; 19222 C; 21909 G; 29264 T; 1 other;

alignment\_scores:

Quality: 4189.00 Length: 813  
Ratio: 5.159 Gaps: 0  
Percent Similarity: 99.877 Percent Identity: 99.754  
alignment\_block:  
US-09-701-711-2 x AAF28550 ..  
Align seg 1/1 to: AAF28550 from: 1 to: 99629  
1 MetArgAsnSerTyrPheLysGlyPheGlnValSerAlaMetThrMetal 17  
|||||  
66133 ATGCGTAATTCATATTTAAAGGTTTTCAGGTGAGTGAATGACATGCAATGGC 66182  
17 aValMetMetValMetSerThrHisAlaGlnAlaAlaAspPheMetAlaA 34  
|||||  
66183 TGTCTATGTTGTAATGTCAACTCATGCCAAGCGGGGATTTATGGCAA 66232  
34 snAspIleThrIleThrGlyLeuGlnArgValThrIleGluSerLeuGln 50  
|||||  
66233 ATGACATTTGCCATCACAGGACTACAGCGAGTACCATTGAAAGCTTACAA 66282  
51 SerValLeuProPheArgLeuGlyGlnValValSerGluAsnGlnLeuAl 67  
|||||  
66283 ACGGTGCTGCCGTTTCGCTTGGGTCAAGTGGTGAGCGAAGCACAGTTGGC 66332  
67 aAspGlyValLysAlaLeuTyrAlaThrGlyAsnPheSerAspValGlnV 84  
|||||  
66333 TGATGGTGTCAAGCACATTATGCAACAGGCAATTTTCAGATGTGCAAG 66382  
84 alTyrHisGlnGluGlyArgIleIleTyrGlnValThrGluArgProLeu 100  
|||||  
66383 TCTATCATCAAGAGGGCGTATCATCTATCAGGTAACCGAAGGCGCTTA 66432  
101 IleAlaGluIleAsnPheGluGlyAsnArgLeuIleProLysGluGlyLe 117  
|||||  
66433 ATCGCTGAGATTAAATTTGAGGGCAATCGCTTAATTTCCAAAAGAGGTCT 66482  
117 uGlnGluGlyLeuLysAsnAlaGlyLeuAlaValGlyGlnProLeuLysG 134  
|||||  
66483 ACAAGAAGGGCTAAAAATGCTGGCTTAGCTGTGGGTCAACCCACTAAAC 66532  
134 InAlaThrValGlnMetIleGluThrGluLeuThrAsnGlnTyrIleSer 150  
|||||  
66533 AAGCCACAGTACAGATGATCGAAACCGAGCTTACCAATCAATATATATCA 66582  
151 GlnGlyTyrTyrAsnThrGluIleThrValLysGlnThrMetLeuAspG 167  
|||||  
66583 CAAGGCTATTATAACCGAAATTTACTGTCAACACAGACGATGCTTGATGG 66632  
167 yAsnArgValLysLeuAspMetThrPheAlaGluGlyLysProAlaArgV 184  
|||||  
66633 TAATCGTGTAAAGCTTATATGACCTTTGCTGAAGGTAAACCTGCACGGG 66682  
184 alValAspIleAsnIleIleGlyAsnGlnHisPheSerAspAlaAspLeu 200  
|||||  
66683 TGGTTGATATTAAATATCATTTGGCAATCAGCATTTTACGCGATGCGATTGG 66732  
201 IleAspValLeuAlaIleLysAspAsnLysIleAsnProLeuSerLysAl 217  
|||||  
66733 ATGTAGTGTCTGGCGATTAAAGGATAATAAAATCAATCTCTCTAAAGC 66782  
217 aAspArgTyrThrGlnGluLysLeuValThrSerLeuGluAsnLeuArgA 234  
|||||  
66783 TGACCGTTATACTCAAGAAAAGCTGGTGACCACTTTAGAGAAATTTGCGTG 66832  
234 laLysTyrLeuAsnAlaGlyPheValArgPheGluIleLysAspAlaLys 250  
|||||  
66833 CTAAATATCTCAATCCAGGGTTTGTGCGCTTTTGAGATTAAAGATGCTAAG 66882  
251 LeuAsnIleAsnGluAspLysAsnArgIlePheValGluIleSerLeuHi 267  
|||||  
66883 CTTAATATTAATGAAGATAAAACCGTACTCTTTGTTGAGATTTTCATTGCA 66932

```

267 sGluGlyGlnTyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuT 284
|||||
66933 TGAAGGTGAGCAATATCGCTTGACAGACACACAGTTTGGGTAATTAA 66982
284 hrTyrThrGlnAlaGluLeuGluAlaLeuLeuLysPheLysAlaGluGlu 300
|||||
66983 CTTTACTACAGCAGAACTTTGAGCGCACTGCTTAAATTCAAAGCAGAAGAA 67032
301 GlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLy 317
|||||
67033 GGGTTTTACAAAGCCATGCTTGACCAACCAACAAACAATATCAGTACCAA 67082
317 sPheGlyAspAspGlyTyrTyrAlaGlnIleArgProValThrArgI 334
|||||
67083 ATTTGGTGACGATGGCTATATTATGCTCAATCCGTCCTGTACACGCA 67132
334 leAsnAspGluSerArgThrValAspValGluTyrTyrIleAspProVal 350
|||||
67133 TTAATGATGAAAGTCGTACGGTTGATGTGGAATATTATTATGACCCCTGTA 67182
351 HisProValTyrValArgArgIleAsnPheThrGlyAsnPheLysThrG1 367
|||||
67183 CACCCTGTCTAATGACCGCTATTAAATTTACAGGTAACTTTAAGACCCA 67232
367 nAspGluValLeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlas 384
|||||
67233 AGATGAAGTACTCCGTCGTGAGATGCACAACTTGAAGGTGCGTTGCAT 67282
384 erAsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhe 400
|||||
67283 CTAATCAAAAAATCAGCTGCTCGTCGACGCTTGATGCGGACTGGGTTT 67332
401 PheLysHisValThrValAspThrArgProValProAsnSerProAspG1 417
|||||
67333 TTTAAACATGTTACCGTTGATCTGCTCCAGTACCCCAACTCACCTGATCA 67382
417 nValAspValAsnPheValValGluGlnProSerGlySerSerThrI 434
|||||
67383 GGTGTGATTAATTTGTGGTGTGAAGAACACACCTTCAGGATCATCAACCA 67432
434 leAlaIaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspVal 450
|||||
67433 TCGCAGCAGGCTACTCTCAAGTGGTGGTAACTTTTCAATTTGATGTT 67482
451 SerGlnAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSe 467
|||||
67483 TCTCAAAATAACTTTATGGGTACAGGTAAGCACCGTCAATGCTTCGTTTC 67532
467 rArgSerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrP 484
|||||
67533 TCGCTCTGAGACCCGAGGTGTATAGTTTGGGTATGACCAACCCACTACT 67582
484 heThrValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThr 500
|||||
67583 TTACCGTAAATGGGCTCTCGCAAGCTTGAGTGGCTACTATCGTAAACCC 67632
501 LysTyrAspAsnLysAsnIleSerAsnTyrValLeuAspSerTyrGlyG1 517
|||||
67633 AAGTATGATAACAAGAACATAGTAATATGTACTTGTATCTTATGTTGG 67682
517 ySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSerPheG 534
|||||
67683 CTCATTAAAGCTATGGATGCAATACCAAGCTTCATGCGCGCTTTTATGGCATT 67732
534 lyLeuAsnAlaAspAsnThrLysLeuHisGlyGlyArgPheMetGlyIle 550
|||||
67733 GTCTGAATGCTGACAATACCAAGCTTCATGCGCGCTTTTATGGCATT 67782
551 SerAsnValLysGlnLeuMetAlaAspGlyGlyLysIleGlnValAspAs 567
|||||
67783 AGTAATGTCAAGCAGCTGATGGCAGATGGTGCAAAATTCAGGTGCGATAA 67832
567 nAsnGlyIleProAspPheLysHisAspTyrThrThrAsnAlaIleL 584

```

```

|||||
67833 TAATGGCATTCCTGATTTAAGCATGATTACACAACCTACAATGCCATTT 67882
584 euGlyTyrAsnTyrSerSerLeuAspArgProValPheProThrGlnGly 600
|||||
67883 TGGGGTGGAAATTAATCAAGTCTAGATCGCCCTGTATTTCCAACCCAAGGC 67932
601 MetSerHisSerValAspLeuThrValIclyPheGlyAspLysThrHisG1 617
|||||
67933 ATGAGTCATTTCTGTAGATTTGACGGTTGGTTTGGTGATAAAACCTCATCA 67982
617 nLysValValTyrGlnGlyAsnIleTyrArgProPheIleLysLysServ 634
|||||
67983 AAAAGTGGTTTATCAAGCAATATCTATCGCCCTATTTACCAAAAAATCAG 68032
634 alLeuArgGlyTyrAlaLysLeuGlyTyrGlyAsnAsnLeuProPheTyr 650
|||||
68033 TCTTGGCTGGATACGCCAAGTTAGGCTATGGCAATAATTTACCATTTTAT 68082
651 GluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667
|||||
68083 GAAATTTCTATCGAGGGCTATGGTTCCGTTCTGCTATGATCAATC 68132
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
|||||
68133 CTCCTTGGGTCCACGCTCACAAAGCCTATTTGACAGCTCGTCTGTTCAAC 68182
684 lnThrThrLeuGlyGluValValGlyGlyAsnAlaLeuAlaThrPheGly 700
|||||
68183 AAACACACTAGGAGAGGTTGTTGGTGGTAAATGCTTTGGCAACTTTCCGC 68232
701 SerGluLeuIleLeuProLeuProPheLysGlyAspTrpIleAspGlnVa 717
|||||
68233 AGTGAGCTGATTTTACCTTTGCCATTTAAAGGTGATTGGATAGATCAGGT 68282
717 lArgProValIlePheIleGluGlyGlyGlnValPheAspThrThrGlyM 734
|||||
68283 GCGTCCAGTGTATTCATTTAGGGCGGTGAGGTTTGTATACAAACAGGTA 68332
734 etAspLysGlnThrIleAspLeuThrGlnPheLysAspProGlnAlaThr 750
|||||
68333 TGGATAACAAACCACTTGATTTAACCCCAATTTAAAGACCCCAAGCAACA 68382
751 AlaGluGlnAsnAlaLysAlaAsnArgProLeuLeuThrGlnAspLys 767
|||||
68383 GCTGAACAAAATGCAAAAGCAGCAATTCGCCGCTACTAACCCCAAGATAA 68432
767 sGlnLeuArgTyrSerAlaGlyValGlyAlaThrTriptyrThrProIleG 784
|||||
68433 ACAGTTGCGTTATAGTCTGTTGGTGGTGGCAACTTGGTATACGCCCATTTG 68482
784 lyProLeuSerIleSerTyrAlaLysProLeuAsnLysLysGlnAsnAsp 800
|||||
68483 GTCCTTATCTATTAGTATGCCAAGCCATTGAATAAAAAACAATAATGAT 68532
801 GlnThrAspThrValGlnPheGlnIleGlySerValPhe 813
|||||
68533 CAGACCGATGAGTACAGTTCCAGATTGGTATGTTGCTTT 68571
seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT: AAS07278
seq_documentation_block:
ID AAS07278 standard; DNA; 2379 BP.
XX
AC AAS07278;
XX
DT 23-OCT-2001 (first entry)
XX
DE Neisseria meningitidis serogroup A antigenic protein DNA.
XX
KW Serogroup B antigen; pharynx; meningitis; septicaemia; mammalian cell;
bacterial infection; baculovirus; yeast; ds.
XX

```

OS Neisseria meningitidis.

XX Key Location/Qualifiers

XX 1..2379

XX CDS

XX /\*tag= a

XX /product= "N. meningitidis serogroup A antigen"

XX sig\_peptide

XX 1..63

XX /\*tag= b

XX mat\_peptide

XX 64..2376

XX /\*tag= c

XX /product= "Mature N. meningitidis serogroup A antigen"

XX WO200138350-A2.

XX 31-MAY-2001.

XX 28-NOV-2000; 2000WO-IB01851.

XX 29-NOV-1999; 99GB-0028197.

XX 09-MAR-2000; 2000GB-0005698.

XX (CHIR-) CHIRON SPA.

XX (STAT-) STATENS INST FOLKEHELSE.

XX Giuliani MM, Pizza M, Rappuoli R, Holst J;

XX WPI; 2001-381289/40.

XX P-PSDB; AAU03958.

XX Novel 85 kDa antigen from Neisseria meningitidis and Neisseria gonorrhoeae, useful in the manufacture of a medicament for treating and preventing Neisserial bacteria infection -

XX Claim 6; Fig 2; 92pp; English.

XX The sequence represents a DNA encoding a Neisseria meningitidis serogroup A 85 kDa antigenic protein. Neisseria meningitidis colonises the pharynx, causing meningitis and, occasionally, septicaemia in the absence of meningitis. This antigenic protein is useful in the manufacture of a medicament for treating or preventing infection due to Neisseria bacteria, such as meningitis and septicaemia. It is also useful as a diagnostic reagent for detecting the presence of Neisseria bacteria or antibodies raised against Neisseria, and as a reagent for raising the antibodies. The Neisserial nucleotide sequences can be expressed in a variety of different expression systems, for example, mammalian cells, baculoviruses, plants, bacteria and yeast.

XX Sequence 2379 BP; 636 A; 755 C; 596 G; 392 T; 0 other;

alignment\_scores:

Quality: 1123.50 Length: 837

Ratio: 2.073 Gaps: 19

Percent Similarity: 64.755 Percent Identity: 32.616

alignment\_block:

US-09-701-711-2 x AAS07278 ..

Align seg 1/1 to: AAS07278 from: 1 to: 2379

10 GlnValSerAlaMetThrMetAlaValMetMetValMetSerThrHisAl 26

4 AAACCTGAAACAGATTGCCCTCCGACCTGATGATGTTGGGCATATCGGCTTT 53

26 aGlnAlaAlaAspPheMetAlaAsnAspIleThrGlyLeuGlnA 43

54 GGCAATTGCCGACTTCAACATCCAGACATCCGTGTGGAAGGCTTGACG 103

43 rgValThrIleGlnSerGlnSerValLeuPropPheArgLeuGln 59

104 GTACCGAGCGGACGACCGTATTCACTACCTGCCGTCGAAGTCGCGGAC 153

60 ValValSerGluAsnGlnLeuAlaAspGlyValValLysAlaLeuTyrAla 76

154 ACCTACAACGACACACGCGAGTGCATCATCAAAAGCCTGTACGCCAC 203

76 rGlyAsnPheSerAspValGlnValTyrHisGlnGluGlyArgIleIle 93

204 CGGTTTCTTTGAGGACGTAGAGTCAAACTGCGGACGGCAGCTTCTGC 253

93 yrGlnValThrGluArgProLeuIleAlaGluIleAsnPheGluGlyAsn 109

254 TGACCGTTATTCGAACGCCCAACCATCGGCTCGCTCAACATCACCGCGCC 303

110 ArgLeuIleProGlyGluGlyLeuGlnGluGlyLeuLysAsnAlaGlyLe 126

304 AAAATGCTCAAAACGACGCCATCAAGAAAAACCTCGAATCGTTGGGCT 353

126 uAlaValGlyGlnProLeuLysGlnAlaThrValGlnMetIleGluThrG 143

354 GCGCGAGTCGCAATACTTTAATCAGGCGACACTCAACAGGCGAGTCGCCG 403

143 luLeuThrAsnGlnTyrIleSerGlnGlyTyrTyrAsnThrGluIleThr 159

404 GCCTGAAAGAAGAAATACCTCGGCGTGGCAAACTCAATATCCAATCACG 453

160 ValLysGlnThrMetLeuAspGlyAsnArgValLysLeuAspMetThrPh 176

454 CCCAAAGTAACCAAACTCGCCCAACCGCTCGACATCGACATCACGAT 503

176 eAlaGluGlyLysProAlaArgValValAspIleAsnIleIleGlyAsnG 193

504 TGACGAGGCGCAATCCGCCAAATCACCGACATCGAATTTGAAGGCAACC 553

193 lnHisPheSerAspAlaAspLeuIleAspValLeuAlaIleLysAspAsn 209

554 AAGTCTATTCCGACCGCAAACTGATCGCGCAGATGTCGCTGACCGAAGGC 603

210 LysIle...AsnProLeuSerLysAlaAspArgTyrThrGlnGluLysLe 225

604 GGCATTGTGACATGGTGTACAGAAAGCAGCGGTTCGACCGCGCAAAAT 653

225 uValThrSerLeuGluAsnLeuArgAlaLysTyrLeuAsnAlaGlyPheV 242

654 CGCCCAAGCATGGAAGGCAACAGTAACCGACTTCTACGAGAACACCGGTACT 703

242 alArgPheGluIleLysAspAlaLysLeuAsnIleAsnGluAspLysAsn 258

704 TCGATTTCCGTATCCTCGATACCGACATCCAAACCAACGAAGACAAAACC 753

259 ArgIlePheValGluIleSerLeuHisGluGlyGluGlnTyrArgPheG 275

754 AGGACAGACCATCAAAATCACCGTCCACGAAGCGGACGTTTCCGCTGGGG 803

275 yGlnThrGlnPheLeuGlyAsnLeuThr...TyrThrGlnAlaGluLeuG 291

804 CAAAGTGTGATTGAAGGCGCACACCAACGAAGTCCCAAGGCGCGAAGTGG 853

291 luAlaLeuLeuLysPheLysAlaGluGluGlyPheSerGlnAlaMetLeu 307

854 AAAAAGTGTGCTGACCATGAAGCCCGCAAAATGGTACGAACGCCAGCAGATG 903

308 GluGlnThrThrAsnAsnIleSerThrLysPheGlyAspAspGlyTyrTy 324

904 ACCGCGGTTTTGGGTGAGATTCAAGAACCGCATGGGCTCGGACAGGCTACGC 953

324 rTyrAlaGlnIleArgProValThrArgIleAsnAspGluSerArgThrV 341

954 ATACCGGAAATCAGCGTACGCGGTCCGCGGACCGCGGGAACCAAAACCG 1003

341 alAspValGluTyrTyrIleAspProValHisProValTyrValArgArg 357

1004 TCGATTTCCGTCCTGCACATCGAACCGCGCGGAAATCTACGTCAACGAA 1053

358 IleAsnPheThrGlyAsnPheLysThrGlnAspGluValLeuArgArgG 374



```
1054 ATCCACATCACCGGCAACAAACAAACCGCGACGAGTCGTGCGCCGCGA 1103
374 uMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnLeuS 391
1104 ATTCGCGCAATGAATCCCGGCTTACGACACCTCCAAAGCTGCAACGCT 1153
391 erArgAlaArgLeuMetArgThrGlyPhePheLysHisValThrValAsp 407
1154 CCAAAGAGCGCGTCGAGCTTTTGGCTACTTCGACAACTGACATTGTGAT 1203
408 ThrArgProValProAsnSerProAspGlnValAspValAsnPheValVa 424
1204 GCCCTCCCGCTTGGCGTAGCCGACAAAGTCGATTGAACATGAGCCT 1253
424 lGluGlnProSerGlySerThrIleAlaAlaGlyTyrSerGlnS 441
1254 GACGAAAGCTCCACCGCTCGCTGCTGACTTGAGCGCGGCTTGGTTTCAGG 1303
441 erGlyGlyValThrPheGlnPheAspValSerGlnAsnAsnPheMetGly 457
1304 ATACCGGCTTGGTCATGTCGCGCGGTATCGCACGAGCAACCTGTCGGT 1353
458 ThrGlyLysHisValAsnAlaSerPheSerArgSerGlnThrArgGluVa 474
1354 ACGGCAAGTCGCGCGCTCGCGCGCTCGGAAAGCAACCCACGCTCAA 1403
474 lTyrSerLeuGlyMetThrAsnProTyrPheThrValAsnGlyValSerG 491
1404 CGGCTCGCTGCTGTTACCGACCGCTACTTCACGGCAGACGGGTACGC. 1452
491 lnSerLeuSerGlyTyr.....TyrArgLysThr.....LysTyr 502
1453 .....CTGGCTAGCATATTTACGAAAGCAAGCTTCGACCCGCGCAA 1494
503 AspAsnLysAsnIleSerAsnTyrValLeuAspSerTyrGlyGlySerLe 519
1495 GCATCGACGAGCTCAACAAATATAAACACACCCACCGCGCGCGCGGT 1544
519 uSerTyrGlyTyrProIleAspGluAsnGlnArgIleSerPheGlyLeuA 536
1545 AAGCATGGTATCCCGTTACCGAATACGACCGCGTCAATTTCCGGGTGG 1594
536 snAlaAspAsnThrLysLeuHisGly..... 544
1595 CGCGCAACACCTGACCGCTCAACACCTTACAAACAAAGCAACCAACGCTAT 1644
545 .....GlyArgPheMetGlyIleSerAsnValLy 554
1645 GCCGACTTATCAGAAATACGGCAACACCGACGCG..... 1680
554 sGlnLeuMetAlaAspGlyGlyLysIleGlnValAspAsnGlyIleP 571
1681 .....GCAGACGCG..... 1689
571 roAspPheLysHisAspTyrThrTyrAsnAlaIleLeuGlyTyrAsn 587
1690 ..ACCTTCAAAGGC.....CTGCTGTACAAAGCACCGCTCGGTGGGG 1731
588 TyrSerSerLeuAspArgProValPheProThrGlnGlyMetSerHisSe 604
1732 CGCAACAGACCGACGGCGTATGCGCGACGCGGCTACCTGACCGG 1781
604 rValAspLeuThrValGlyPhe...GlyAspLys...ThrHisGlnLysV 619
1782 CGTAATGCGCAATTCGCCCTGCGCGCAGCAAACTGCAATACTACTCCG 1831
619 alValTyrGlnGlyAsnIleTyrArgProPheIleLysLysSerValLeu 635
1832 CCACCCCAACCAACCTGTTCTTCCCTTAAGCAAAACCTTCACGCTG 1881
636 Arg.....GlyTyrAlaLysLeuGlyTyrGly.....As 645
1882 ATGCTCGGCGCGAAGTCGGCATTTGGGCG.....GGCTACGCGCAGAACCA 1928
```

```
645 nAsnLeuProPheTyrGluAsnPheTyrAlaGlyGlyTyrGlySerVala 662
1929 AGAAATCCCTCTTTGAAAACITCTACGGCGGCGGCTGGTTTCGGTGC 1978
662 rgGlyTyrAspGlnSerSerLeuGlyProArgSerGlnAlaTyrLeuThr 678
1979 CGCGCTACGAAAGCGCGCTCGCGCGGAA..... 2010
679 AlaArgArgGlyGlnGlnThrThrLeuGlyGluVal.....GlyG 693
2011 .....GTGTATGAGCAATACGGCGAAATAATCAGCTACGCGCG 2048
693 yAsnAlaLeuAlaThrPheGlySerGluLeuIleLeuProLeuProPheL 710
2049 CAACAAAAGCCCAACCTCTCCGCGAGCTCTTCCCGATGCCGCGTG 2098
710 ysGlyAspTyrPheAspGlnValArgProValIlePheIleGluGlyGly 726
2099 CGAAAGAC...GCACGACCGCTCGCTGAGCTGTTTTCGCGACGACG 2145
727 GlnValPheAspThrThrGlyMetAspLysGlnThrIleAspLeuThrG 743
2146 AGCGTGGG.....GACGGCAGAACCTATACCGCGCGCGA 2180
743 nPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnA 760
2181 AACCGGTAAACAACAATCGGTTTACTCGGAAACGCGCATAATCC.... 2226
760 rgProLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGly 776
2227 .....ACCTTTACCAACAAATTCGCGTATTCGCGCGCGCGCG 2265
777 AlaThrTyrTyrThrProIleGlyProLeuSerIleSerTyrAlaLysPr 793
2266 GTTACCTGGCTCGCTCGCTTGGTCCGATGAAATCAGTACGCTACCC 2315
793 oLeuAsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGlnIleG 810
2316 GCTGAAGAAAAACCGGAAGACGAAATCCACGCTTCCAAATTCAGCTCG 2365
810 lySerValPhe 813
2366 GCACGACGTT 2376
seq name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH42129
seq documentation_block:
ID AAH42129 standard; DNA; 2379 BP.
XX
AC AAH42129;
XX
DT 17-SEP-2001 (first entry)
XX
DE Nucleotide sequence of a Neisseria gonorrhoea protein.
XX
KW Serogroup B protein; outer membrane protein; Neisserial infection;
KW vaccine; ss.
XX
OS Neisseria gonorrhoea.
XX
FH Key Location/Qualifiers
CDS 1..2379
FT /tag= a
FT /product= "Neisseria serogroup B protein"
FT sig_peptide 1..63
FT /tag= b
FT mat_peptide 64..2376
FT /tag= c
XX
PN WO200152885-A1.
XX
PD 26-JUL-2001.
```









1404 CGGCTCGCTGCTGTTACTACCGCTACTTACGGCAGACGGGGTCAGCC 1453  
491 InSerLeuSerGlyTyrTyrArg.....LysThrLysTyrAspAsnLys 505  
1454 TGGGCTACGATGTTTACGGAAAGCCCTCGACCCGCGCAAGCATCGACC 1503  
506 AsnLeuSerAsnTyrValLeuAspSerTyrGlyGlySerLeuSerTyrG1 522  
1504 AGCATCAACAATATAAAACACCAACCGCAGCGCAGGCATCCGCATGAG 1553  
522 yTyrProLeuAspGlnAsnGlnArgIleSerPheGlyLeuAsnAlaAspA 539  
1554 CGTCCCTGTTACCAATACGACCGCTGTAATTCGGTTGGTGGCAGAAC 1603  
539 snThrLysLeuHisGlyGlyArgPheMetGlyIleSerAsnValLysGln 555  
1604 ACCTGACCGCTCAAC.....ACCTACAACAAGCCGCCA 1638  
556 LeuMetAlaAspGlyGlyIleGlnValAspAsnAsnGlyIleProAs 572  
1639 CACTATGCGGACTTATCAAGAATACGGCAAAACCGACGCG...ACAGA 1685  
572 pPheLysHisAspTyrThrThrTyrAsnAlaIleLeuGlyTrpAsnTyrS 589  
1686 CGCAGAGCTTCAAAGGCTGGCTGTACAAGGTACCGTGGCTGGGGCGCA 1735  
589 erSerLeuAspArgProValPheProThrGlnGlyMetSerHisSerVal 605  
1736 ACAAAACCGCAGCGGCTTATGGCGGCGCGCGGCTACCTACGCGGGGTG 1785  
606 AspLeuThrValGlyPhe...GlyAspLys...ThrHisGlnLysValVa 620  
1786 AACGCCGAATCGCCCTCGCTGGCGCAAACTGCAATACTACTCGGCCAC 1835  
620 lTyrGlnGlyAsnLeTyrArgProPheIleLeLysLysSerValLeuArg. 636  
1836 CCACAACCAACACCTGGTCTTCCCTCCCTGAGCAAAACCTTACGCTGATGC 1885  
637 .....GlyTyrAlaLysLeuGlyTyrGly.....AsnAsn 646  
1886 TCGCGGGGAAGTCGGCATTCGGGCG...GGCTACGGCAGAACCAAGAA 1932  
647 LeuProPheTyrGluAsnPheTyrAlaGlyGlyTyrGlySerValArgG1 663  
1933 ATCCCTCTTTTGAATACTTCTACGGCGCGGCTGGTTCGGTTCGGCGCG 1982  
663 yTyrAspGlnSerSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaA 680  
1983 ATACGAAGCGGCGCTCGGTCGCGAAA..... 2010  
680 rGArgGlyGlnGlnThrThrLeuGlyGluValVal.....GlyGlyAsn 694  
2011 .....GTCTATGACCAATACCGCGCAAAATAACAGTACGGCGGCAAC 2052  
695 AlaLeuAlaThrPheGlySerGluLeuLeuLeuProLeuProPheLysG1 711  
2053 AAAAAAGCCAACGTCTCGCGGAGCTGCTCTTCCCGATGCCCGCGCGAA 2102  
711 yAspTrpIleAspGlnValArgProValIlePheIleGluGlyGlnV 728  
2103 AGAC...GCGCGCACCGTCCGCGCTGAGCCTGTTGCCGACGCGGACGCG 2149  
728 alPheAspThrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPhe 744  
2150 TGTGGGACGGCAAAACCTTACGACCAACACAGCAGTTCGGCGACC..... 2193  
745 LysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgPr 761  
2194 .....GCGCGGAGGGTTCAAAACATTTACGCGCGCGGCAATACCCA 2234  
761 oLeuLeuThrGlnAspGlnLeuArgTyrSerAlaGlyValGlyAlat 778

2235 TAAATCCACCTTTACCAACGAATTGCGCTATTCGCCGCGCGCGGGTTA 2284  
778 hrTptTyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeu 794  
2285 CCTGGCTCTCGCCTTTAGGCCCGCATGAATTCAGTACGCTACCCGCTG 2334  
795 AsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGlnIleGlySe 811  
2335 AAGAAAAACCGGAAGCAAGAAATCCACGCTTCCAATTCCAACTCCGCAC 2384  
811 rValphe 813  
2385 GACGTTT 2391  
seq\_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH42128  
seq\_documentation\_block:  
ID AAH42128 standard; DNA; 2394 BP.  
XX  
AC AAH42128;  
XX  
DT 17-SEP-2001 (first entry)  
XX  
DE Nucleotide sequence of a Neisseria serogroup B protein.  
XX  
KW Serogroup B protein; outer membrane protein; Neisserial infection;  
KW vaccine; ss.  
XX  
OS Neisseria meningitidis.  
XX  
FH Key Location/Qualifiers  
CDS 1..2394  
FT /\*tag= a  
FT /\*product= "Neisseria serogroup B protein"  
FT sig\_peptide 1..63  
FT /\*tag= b  
FT mat\_peptide 64..2391  
FT /\*tag= c  
XX  
PN WO200152885-A1.  
XX  
PD 26-JUL-2001.  
XX  
PF 17-JAN-2001; 2001WO-IB00166.  
XX  
PR 17-JAN-2000; 2000GB-0001067.  
PR 09-MAR-2000; 2000GB-0005699.  
XX  
PA (CHIR-) CHIRON SPA.  
XX  
PI Pizza M, Rappuoli R, Giuliani M;  
XX  
WPI: 2001-451895/48.  
DR P-PSDB; AAB84744.  
XX  
Composition for treating or preventing infection to, detecting, or for raising antibodies against Neisserial bacteria, comprises an N. meningitidis serogroup B outer membrane preparation and an immunogenic component -  
XX  
Disclosure; Page 58-59; 83pp; English.  
XX  
The present sequence encodes a Neisseria serogroup B protein. The protein is used to produce the compositions of the invention. The specification describes a composition, comprising a Neisseria meningitidis serogroup B outer membrane preparation and an immunogenic component. The immunogenic component is protein disclosed in WO99/57280, WO99/36544, WO99/24578, WO99/66791, WO99/28273, WO96/29412, WO95/03413, WO99/31132, WO99/58683, WO99/55873, and/or N. meningitidis protein PorA, TbpA, TbpB, PilC, Opa, or Omp85. The composition is used for making a medicament for treating or preventing infection due to Neisserial bacteria; a diagnostic reagent for detecting the presence of Neisserial bacteria or of antibodies raised against Neisserial bacteria; and/or



```

1604 ACCTGACCGCTAAC.....ACCTACAAACAAAGCCCAAA 1638
556 LeuMetAlaaspGlyGlyLysLeuValAspAsnGlyIleProAs 572
      |||||
1639 CACTATGCCGACTTATCAGAAATACGGCAAAACGACGGC...ACAGA 1685
572 pPheLysHisAspTyrThrThrTyrAsnAlaIleLeuGlyTyrAsnTyrS 589
      |   :   :   :   :   :   :   :   :   :   :   :
1686 CGGCAGCTTCAAAGCGTGTGTACAAAGGTACCGTGGTGGGGCGCA 1735
589 erSerLeuAspArgProValPheProThrGlnGlyMetSerHisSerVal 605
      |   :   :   :   :   :   :   :   :   :   :   :
1736 ACAAACCGCAGCGCGCTTATGCGCGACGCGCGCTACCTACGCGCGTG 1785
606 AspLeuThrValGlyPhe...GlyAspLys...ThrHisGlnLysValVa 620
      :   :   :   :   :   :   :   :   :   :   :
1786 AACCGCGAATCGCCCTGCGTGGCAGCAAACTGCAATACTACTCGGCAC 1835
620 lTyrGlnGlyAsnIleTyrArgPropheIleLysLysSerValLeuArg. 636
      :   :   :   :   :   :   :   :   :   :   :
1836 CCACAACCAAACTGTTCTTCCCTCGAGCAAAACCTTACGCTGATGC 1885
637 .....GlyTyrAlaLysLeuGlyTyrGly.....AsnAsn 646
      ||| |||
1886 TCGCGCGGAGTGGCATTCGGGCG...GGCTACGGCAGAACCAAGAA 1932
647 LeuProPheTyrGluAsnPheTyrAlaGlyTyrGlySerValArgG1 663
      |||||
1933 ATCCCTTCCTTGAATACTTCTACGGCGCGGCTGGTTCGGTTCGGCGG 1982
663 yTyrAspGlnSerSerLeuGlyProArgSerGlnAlaTyrLeuThrAla 680
      |||||
1983 ATACGAAGCGCAGCTCGTCCGAA..... 2010
680 rGArgGlyGlnGlnThrThrLeuGlyGluValVal.....GlyGlyAsn 694
      |||||
2011 .....GTCTATGACGAATACGGCGCAAAATACAGTACGGCGGCAAC 2052
695 AlaLeuAlaThrPheGlySerGluLeuLeuProLeuProPheLysG1 711
      |||||
2053 AAAAAGCAACACGCTCCGCGAGCTGCTCTCCGATGCCGCGCGCAA 2102
711 YAspTrpIleAspGlnValArgProValIlePheIleGluGlyGlnV 728
      |||
2103 AGAC...GCCGCGACCTGCGCTGAGCCCTGTTCGCGCAGCGACGG 2149
728 alPheAspThrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPhe 744
      |||||
2150 TGTGGGACGCAAAACCTACGACGACACAGCAGTTCGCGGACC..... 2193
745 LysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAsnArgPr 761
      :   :   :   :   :   :   :   :   :   :   :
2194 .....GGCGCAGGGTTCAAACATTTACGGCGCGCGCAATACCCA 2234
761 oLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAla 778
      |||
2235 TAATCCACTTTACCAACGAATTCGCTATTCGCGCGCGCGCGGTTA 2284
778 hrTrpTyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeu 794
      |||||
2285 CCGTGCCTCGCTTTAGGCGCGATGAATTCAGCTACGCTACCGCTG 2334
795 AsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGlnIleGly 811
      :   :   :   :   :   :   :   :   :   :   :
2335 AAAAAAACCAGGAAGCAAAATCCACGCTTCCAAATTCCAACTCGGCAC 2384
811 rValPhe 813
      :   :   :   :   :
2385 GAGCTTC 2391

```

```

seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AAA81478
seq_documentation_block:

```

```

ID AAA81478 standard; DNA; 52253 BP.
XX
AC AAA81478;
XX
DT 04-DEC-2000 (first entry)
XX
DE N. meningitidis partial DNA sequence gnm_26 SEQ ID NO:26.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW Meningococcus B; MenB; ds.
XX
OS Neisseria meningitidis.
XX
PN WO200022430-A2.
XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-US23573.
XX
PR 09-OCT-1998; 98US-0103794.
PR 30-APR-1999; 99US-0132068.
XX
PA (CHIR ) CHIRON CORP.
XX
PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
PI Massignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
PI Rappuoli R, Pizza M;
XX
DR WPI; 2000-318079/27.
XX
PT Isolated nucleotide sequences of Neisseria meningitidis which can be
PT used in the diagnosis and treatment of N. meningitidis infection and
PT other Neisserial infections, for example, N.gonorrhoea.
XX
PS Claim 7; Page 532-547; 1760pp; English.
XX
CC The present invention describes methods of obtaining immunogenic
CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
CC represent specifically claimed Neisseria meningitidis genomic DNA
CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to
CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC sequences, which are all used in the exemplification of the present
CC invention. The nucleic acid sequences, protein sequences, and antibodies
CC against them, can be used in the manufacture of a composition. The
CC composition can be used as a medicament (or in the manufacture of a
CC medicament) for treating, preventing or diagnosing infection due to
CC Neisserial bacteria. For example, some of the identified proteins could
CC be components of vaccines against Meningococcus B; against all serotypes;
CC and/or against all pathogenic Neissariae. Identification of sequences
CC from the bacterium will also facilitate production of biological probes,
CC particularly organism-specific probes. Attempts to make efficacious
CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC Multivalent vaccines have also been tried but none have successfully
CC overcome antigenic variability. The provision of further, complete
CC sequences may provide an opportunity to identify secreted or surface
CC exposed proteins that may be presumed targets for the immune system and
CC which are not antigenically variable or at least more conserved than
CC other more variable regions.
XX
SQ Sequence 52253 BP; 12877 A; 15510 C; 13103 G; 10761 T; 2 other;

```

```

alignment_scores:
  Quality: 1115.00      Length: 819
  Ratio: 2.042         Gaps: 14
  Percent Similarity: 66.667   Percent Identity: 32.479

alignment_block:
US-09-701-711-2 x AAA81478

```

Align seg 1/1 to: AAA81478 from: 1 to: 52253

10	GlnValSerAlaMetThrMetAlaValMetMetValMetSerThrHisAl	26
.....	.....	.....
47301	AAACTGAAACAGATTGCTTCGGCACTGATGATGTGGCATATCGCCTTT	47350
26	aglnAlaAlaAspPheMetAlaAsnAspIleThrIleThrGlyLeuGlnA	43
.....	.....	.....
47351	GGCACTTGGCACTTCAACATCCAAGACATCCGGCTCGAAGGCTTGACG	47400
43	rgValThrIleGluSerSerValLeuProPheArgLeuGlyGln	59
.....	.....	.....
47401	GTACCGAGCGGATACCGTATTCACTACCTGCGCGTCAAAAGTCGGGAC	47450
60	ValValSerGluAsnGlnLeuAlaAspGlyValLysAlaLeuTyrAlaTh	76
.....	.....	.....
47451	ACCTACAACGACACACACGCGAGTGCCATCATCAAAAGCCGTGACGCCAC	47500
76	rgLysAsnPheSerAspValGlnValTyrHisGlnGluGlyArgIleIle	93
.....	.....	.....
47501	CGGTTCCTTGGACGACGTACGCGTCTGAAACTTGGCGAGCGGACGCTCTGC	47550
93	yrGlnValThrGluArgProLeuIleAlaGluIleAsnPheGluGlyAsn	109
.....	.....	.....
47551	TGACCGTTATCGAAGCGCCACCATCCGGCTCGCTCAACATCACCGGCGCA	47600
110	ArgLeuIleProLysGluGlyLeuGlnGlnGluGlyLeuLysAsnAlaGlyLe	126
.....	.....	.....
47601	AAAAAGTGTCAAAACGAGCGCATTAAGAAAAACCTCGAATCGTTTCGGGCT	47650
126	uAlaValGlyGlnProLeuLysGlnAlaThrValGlnMetIleGluThrG	143
.....	.....	.....
47651	GGCGCAGTCCCAATACTTTAATCAGCGCACATCAATCAGCGACATCGCGC	47700
143	luLeuThrAsnGlnTyrIleSerGlnGlyTyrTyrAsnThrGluIleThr	159
.....	.....	.....
47701	GCCTGAAGAAGAAATACCTCGGGCGCGGCAAACTCAATATCCAATCAG	47750
160	ValLysGlnThrMetLeuAspGlyAsnArgValLysLeuAspMetThrPh	176
.....	.....	.....
47751	CCCCAAAGTACCAAACTCGCCGCAACCGGCTCGACATCGACATCAGAT	47800
176	eAlaGluGlyLysProAlaArgValValAspIleAsnIleIleGlyAsnG	193
.....	.....	.....
47801	TGACGAGGGCAATCGGCCAAAATCACCGACATCGAATTTGAAGCGCAACC	47850
193	lnHisPheSerAspAlaAspLeuIleAspValLeuAlaIleLysAspAsn	209
.....	.....	.....
47851	AAGTCATTCCGACCGCAAACTGATCGGCAAAATGTCCTCACCAGAGGC	47900
210	LysIle...AsnProLeuSerLysAlaAspArgTyrThrGlnGluLysLe	225
.....	.....	.....
47901	GGCATTGTGGACATGCTGACACGAAGCAACCAATTCACAGCAGCAGAATT	47950
225	uValThrSerLeuGluAsnLeuArgAlaLysTyrLeuAsnAlaGlyPheV	242
.....	.....	.....
47951	TGCCCAAGATATGGAAGAAGTACCGACTTCTACCAAAATTAACGGCTACT	48000
242	aArgPheGluIleLysAspAlaLysLeuAsnIleAsnGluAspLysAsn	258
.....	.....	.....
48001	TCGATTTCCTCATCTCGATACCGACATCCAAACCAACGAAGACAAACC	48050
259	ArgIlePheValGluIleSerLeuHisGluGlyGluGlnTyrArgPheGl	275
.....	.....	.....
48051	AACGACACCATCAAAATCACCGTCCAGACGAGCGGACGTTTCCGCTGGGG	48100
275	yGlnThrGlnPheLeuGlyAsnLeuThr...TyrThrGlnAlaGluLeuG	291
.....	.....	.....
48101	CAAAAGTCTCCATCGAGGCGACACCAACGAAGTCCCCCAAGCCGCACTGG	48150
291	luAlaLeuLysPheLysAlaGluGlyPheSerGlnAlaMetLeu	307
.....	.....	.....

48151 AAAAACTGCTGACCATGAAGCCCGGCAAAATGGTTCAGAACGCCAGCAGATG 48200

308 GluGlnThrThrAsnAsnIleSerThrLysPheGlyAspSpGlyTyrrTy 324  
::: :::::|||||:::|||||:::|||||:::|||||:::|||||  
48201 ACCGCGTTTGGGTGAGATTCAAGACCGCATGGCTCGGCAGCTACGC 48250

324 rTyrAlaGlnIleArgProValThrArgIleAsnAspGluSerArgThrV 341  
|||:::||||| |||:::||||| |||:::||||| |||:::|||||  
48251 ATACAGCGAATAATCACGCTACAGCCGCTCCGCAAGCGCTCAAACCAAAACCG 48300

341 alaspValcIuTyrrTyrtleaspProValHisProValTyrrValArgarg 357  
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
48301 TCGATTTCGTCTGCATCATCAAGACGGCGCGGAAAATCTAGCTCAACGAA 48350

358 IleAsnPheThrGlyAsnPheLysThrGlnAspGluValLeuArgArgGl 374  
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
48351 ATPACACATCACGGGCAACAACAAACCCGCGACGAAGTCGTCGCCGCTGA 48400

374 uMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnLeuS 391  
|||:::||||| |||:::||||| |||:::||||| |||:::||||| |||:::|||||  
48401 ATTACGCCAAATGGAATCCGACACCTTAGCAGACCTCCAAGCTGCAACGTT 48450

391 erArgAlaArgLeuMetArgThrGlyPhePheLysHisValThrValAsp 407  
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
48451 CCAAAAGCGCGGTGCGAGTGTGGTGCTACTTCGACAATGTCCAGTTTGTAT 48500

408 ThrArgProValProAsnSerProAspGlnValAspValAsnPheValVa 424  
::: |||||::: :::::|||||::: |||||::: |||||::: |||||::: |||||:::  
48501 GGTGTCCGCTTGGCGGCACGCCGCAAAATCGATTTGAACATGACTCT 48550

424 IolugLuGlnProSerGlySerSerThrIleAlaAlaGlyTyrSerGlnS 441  
|||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::  
48551 GACCGAACGTTCCACCGGTTCCCTGGATTGAGCGCGGTTGGGTTCAAG 48600

441 erGlycylValThrPheGlnPheaspaValSerGlnAsnAsnPheMetGly 457  
:: |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::  
48601 ATACCGGGTGTGTCATGTCGCGAGCGGCTTCCCAGAACAACGCTGTCGGT 48650

458 ThrGlyLysHisValAsnAlaSerPheSerArgSerGluThrArgGluVa 474  
||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
48651 ACGGCAGTGGCGCGCACTGGCGCCTTCAGGAGCAAAACCCACGCTTAA 48700

474 ltyrSerLeuGlyMetThrAsnProtyrPheThrValAsnGlyValserG 491  
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
48701 CGGCTCGCTGCTGTTACTGACCGCTACTTCAGCGCACACGGGTCAGCC 48750

491 InSerLeuSerGlyTyrrTyrrarg..... LysThrLysTyrrAspAsnLys 505  
::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::  
48751 TGGGCTACGATGTTTACGGAAAAGCGCTTCGACCCGCGGAAAGCATCGACC 48800

506 AsnIleSerAsnTyrrValLeuAspSerTyrrGlyGlySerLeuSerTyrrGl 522  
:::|||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::  
48801 AGCATCAACAAATATAAACCACCGCAGCGCGGCGATCCCGCATGAG 48850

522 ytyrProIleAspGluAsnGlnArgIleSerPheGlyLeuAsnAlaAspA 539  
::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::  
48851 CGTGCCCTGTTACCGAAATACGACCGGCTCAATTCGTTTGGTGGCAGAAC 48900

539 snThrLysLeuHiscylGlyArgPheMetGlyIleSerAsnValLysGln 555  
:: |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::  
48901 ACCTGACCGCTCAAC.....ACCTACAACAAAGCGCCCAAA 48935

556 LeuMetAlaAspGlyLysIleGlnValAspAsnAsnGlyIleProAs 572  
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
48936 CACTATGCGCACTTTATCAAGAAATACGGCAAAACCGACGGC...ACAGA 48982

572 pPheLysHisAspyrThrThrTyrrAsnAlaIleLeuGlyTrpAsnTyrs 589  
|||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::  
48983 CGGCAGCTTCAAGGCTGGCTCTACAAAGGTACCTCGCTCGCTGGGGCGCA 49032

589 erSerLeuAspArgProValPheProThrGlnGlyMetSerHisSerVal 605  
::::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::  
49033 ACAAAACCGCAGACGCGTATTATGGCGACGCGGCTACTTCACGCGGCTG 49082

```
606 AspLeuThrValGlyPhe...GlyAspLys...ThrHisGlnLysValva 620
    ::::
    ::::
49083 AACGCCGAATCGCCTCGCTGGCAGCAACTGCAATACTACTCGCCAC 49132
    ::::
    ::::
620 ltyrGlnGlyAsnIleTyArgProPheIleLysSerValLeuArg. 636
    ::::
    ::::
49133 CCACAACCAAACTGGTTCCTCCCGCTGAGCAAACTTCACGCTGATGC 49182
    ::::
    ::::
637 .....GlyTyAlaLysLeuGlyTyArgLys.....AsnAsn 646
    ::::
    ::::
49183 TCGCGCGCAAGTCGGCATTCGGGC...GGCTACGGCAGAACCAAGAA 49229
    ::::
    ::::
647 LeuProPheTyGluAsnPheTyAlaGlyGlyTyArgLysSerValArgG1 663
    ::::
    ::::
49230 ATCCCTCTCTTGTGAAACTTCTACGGCGCGCTGGTTCGGTTCGGCGG 49279
    ::::
    ::::
663 yTyAspGlnSerSerLeuGlyProArgSerGlnAlaTyLeuThrAlaA 680
    ::::
    ::::
49280 ATACGAAGCGCGACGCTCGGTCCGAAA..... 49307
    ::::
    ::::
680 rGArgGlyGlnThrThrLeuGlyGluValVal.....GlyGlyAsn 694
    ::::
    ::::
49308 .....GTCTATCAGCAATACGGCGAATAAATCAGCTACGGCGCAAC 49349
    ::::
    ::::
695 AlaLeuAlaThrPheGlySerGluLeuLeuLeuProLeuProPheLysG1 711
    ::::
    ::::
49350 AAAAAGCAACAGCTCTCGCGCAGCTGCTCTCCCGATGCCGCGCGAA 49399
    ::::
    ::::
711 yAspTrpIleAspGlnValArgProValIlePheIleLeuGlyGlnV 728
    ::::
    ::::
49400 AGAC....CGCGCGACGTCGCCCTGAGCGCTGTTTCCGACGACGCGAGC 49446
    ::::
    ::::
728 alPheAspThrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPhe 744
    ::::
    ::::
49447 TGTGGCGCGCAAAACCTACGACGACACACAGCTTCGCGCACC..... 49490
    ::::
    ::::
745 LysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgPr 761
    ::::
    ::::
49491 .....GCGCGCAGCGGTTCAAAACATTTACGCGCGCGCAATACCCA 49531
    ::::
    ::::
761 oLeuLeuThrGlnAspLysGlnLeuArgTyTrSerAlaGlyValGlyAlat 778
    ::::
    ::::
49532 TAATCCACCTTTACCAACGAAATGCGCTATTCGCGCGCGCGCGTTA 49581
    ::::
    ::::
778 hrTpTyThrProIleGlyProLeuSerIleSerTyAlaLysProLeu 794
    ::::
    ::::
49582 CTGGCTCTCGCTTTAGCGCGCATGAAATTCAGCTACGCTACCGCGCTG 49631
    ::::
    ::::
795 AsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGlnIleGlySe 811
    ::::
    ::::
49632 AAGAAAACCGGAAGACGAAATCCACGCTTCCCAATTCACCTCGGCAC 49681
    ::::
    ::::
811 rValPhe 813
    ::::
49682 GACGTC 49688

seq_name: /STDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAF21544
seq_documentation_block:
ID AAF21544 standard; DNA; 349980 BP.
XX
AC AAF21544;
XX
DT 13-MAR-2001 (first entry)
XX
DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:1.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;
XX
OS Neisseria meningitidis.
```





```

178874 ATCCCCCTTTTGAATACTTCTACGGCGGGCGCTGGGTGGTGGTGGCGG 178825
663 yTyrAspGlnSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaA 680
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
178824 ATACGAAGCGGCACGCTCGGTCCGAAA..... 178797
680 rGArgGlyGlnGlnThrThrLeuGlyGluValVal.....GlyGlyAsn 694
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
178796 .....GTCTATGAGGAATACGCGGAAATAATCAGCTACGCGGCAAC 178755
695 AlaLeuAlaThrPheGlySerGluLeuLeuLeuProLeuProPheLysG1 711
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
178754 AAAAAACCCACAGCTCCGCGGAGCTCTCTCCGATGCCGCGCGGAA 178705
711 yAspTrpIleAspGlnValArgProValIlePheIleGluGlyGlyGlnV 728
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
178704 AGAC...GCGGCGACCGTCCGCGCTGAGCCTGTTGCGGACGCGGACG 178658
728 alPheAspThrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPhe 744
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
178657 TGTGGGACGCGCAAAACCTACGACGACACACAGCAGTTCGCGGACC..... 178614
745 LysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaLysAlaAsnArgPr 761
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
178613 .....GCGGCGAGGTTCAAAACATTTACGCGCGCGGCATATACCA 178573
761 oLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlar 778
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
178572 TAAATCCACCTTTTACCAAGAAATTCGCTATTCCGCGCGGCGCGGTTA 178523
778 hrTpTyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeu 794
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
178522 CCTGGCTCTCGCCTTTAGGCCGATGAATTCAGCTACGCGCTACCCGCTG 178473
795 AsnLysGlnAsnAspGlnThrAspThrValGlnPheGlnIleGlySe 811
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
178472 AAAAAAAACCGGAAGAGAAATCCACGCTTCCATTCACATCCGAC 178423
811 rValPhe 813
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
178422 GACGTT 178416

```

seq\_name: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AAS07279

seq\_documentation\_block:

ID AAS07279 standard; DNA; 2394 BP.

XX AAS07279;

XX 23-OCT-2001 (first entry)

XX Neisseria gonorrhoeae antigenic protein DNA.

XX Antigenic protein; pharynx; meningitis; septicaemia; mammalian cell;  
 KW bacterial infection; baculovirus; yeast; ds; Neisseria meningitidis.

XX Neisseria gonorrhoeae.

XX Key Location/Qualifiers

FT CDS 1..2394

FT /tag= a

FT /product= "N. gonorrhoeae antigen"

FT sig\_peptide 1..63

FT FT 64..2391

FT mat\_peptide

FT /tag= c

FT /product= "Mature N. gonorrhoeae antigen"

XX W0200138350-A2.

PN 31-MAY-2001.-

XX 28-NOV-2000; 2000WO-IB01851.

PF

```

XX 29-NOV-1999; 99GB-0028197.
PR 09-MAR-2000; 2000GB-0005698.
XX
XX (CHIR-) CHIRON SPA.
PA (STAT-) STATENS INST FOLKEHELSE.
XX
XX Giuliani MM, Pizza M, Rappuoli R, Holst J;
PI
XX WPI; 2001-381289/40.
DR P-PSDB; AAU03959.
XX
XX Novel 85 kDa antigen from Neisseria meningitidis and Neisseria
PT gonorrhoeae, useful in the manufacture of a medicament for treating and
PT preventing Neisserial bacteria infection -
XX
XX Claim 6; Fig 2; 92pp; English.
XX
XX The sequence represents a DNA encoding a Neisseria gonorrhoeae 85 kDa
CC antigenic protein. N. gonorrhoeae is closely related to N. meningitidis,
CC which colonises the pharynx, causing meningitis and, occasionally,
CC septicaemia in the absence of meningitis. This antigenic protein is
CC useful in the manufacture of a medicament for treating or preventing
CC infection due to Neisseria bacteria, such as meningitis and septicaemia.
CC It is also useful as a diagnostic reagent for detecting the presence of
CC Neisseria bacteria or antibodies raised against Neisseria, and as a
CC reagent for raising the antibodies. The Neisserial nucleotide sequences
CC can be expressed in a variety of different expression systems, for
CC example, mammalian cells, baculoviruses, plants, bacteria and yeast.
XX
XX Sequence 2394 BP; 655 A; 758 C; 586 G; 395 T; 0 other;

```

alignment\_scores:

Quality: 1111.00 Length: 819

Ratio: 2.035 Gaps: 14

Percent Similarity: 66.667 Percent Identity: 32.357

alignment\_block:

US-09-701-711-2 x AAS07279 ..

Align seg 1/1 to: AAS07279 from: 1 to: 2394

```

10 GlnValSerAlaMetThrMetAlaValMetMetValMetSerThrHisAl 26
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
4 AAACCTGAACACAGATGCTTCGCGACTGATGCTTGGGCATATCGCCCTT 53
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
26 aGlnAlaAlaAspPheMetAlaAsnAspIleThrIleThrGlyLeuGlnA 43
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
54 GGCACCTTGGCCGACTTCACCATCCAGACATCCGGCTCGAAGGCTTGACAG 103
43 rgValThrIleGluSerLeuGlnSerValLeuProPheArgLeuGlyGln 59
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
104 GTACCGAGCGGAGTACCGTATTCACTACCTGCCGTCAAAAGTCGGCGAC 153
60 ValValSerGluAsnGlnLeuAlaAspGlyValLysAlaLeuTyrAlaTh 76
:::||||:||||:||||:||||:||||:||||:||||:||||:||||:
154 ACCTACACACGACACACGCGGAGTCCCATCATCAAAAGCCTGTACGCCAC 203
76 rGlyAsnPheSerAspValGlnValTyrHisGlnGluGlyArgIleIleT 93
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
204 CGGTTTCTTTGACGACGTTACCGGTGCGGAACCTCGGACGGGCGCTCCG 253
93 yrGlnValThrGluArgProLeuIleAlaGluIleAsnPheGluGlyAsn 109
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
254 TGACCGTTATCGAACGCCGCCACCATCGCTCGCTCAACATCACCAGCGCA 303
110 ArgLeuIleProLysGluGlyLeuGlnGluGlyLeuLysAsnAlaGlyLe 126
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
304 AAAATGCTGCAAAACGACGCGCATTAAGAAAAACCTCGAATCGTTCGGCT 353
126 uAlaValGlyGlnProLeuLysGlnAlaThrValGlnMetIleGluThrG 143
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

354 GCGCGAGTCGCAATACTTTAATCAGGCGCACTCAATCAGCGCATCGCGG 403

143 luLeuThrAsnGlnTyrIleSerGlnGlyTyrTrpAsnThrGluIleThr 159  
||| :|||:||||:||||:||||| ||| :|||:|||||

404 GCCTGAAGAAGAAATACCTCGGGCGCGCAAACTCAATATCCAATCAGG 453

160 ValLysGlnThrMetLeuAspGlyAsnArgValLysLeuAspMetThrPh 176  
||| ||| ||| ||||| ||| :|||:||||| :|||:|||||

454 CCCAAAGTAACCAACTCGCCCGCAAGCGCTCGACATCGACATCAGAT 503

176 eAlaGluGlyProAlaArgValValAspIleAsnIleIleGlyAsnG 193  
||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||

504 TGACGAGGGCAATCCGGCAAAATCACCAGATCGAATTTGAAGGCAACC 553

193 InHisPheSerAspAlaLeuIleAspValLeuAlaIleLysAspAsn 209  
||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||

554 AAGTCTATTCGACCGCAAACTGATCGCGCAGATGTCGCTGACCGAAGG 603

210 LysIle...AsnProLeuSerLysAlaAspArgTyrThrGlnGluLysLe 225  
||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||

604 GGCATTTGGACATGCTGACAGGAACCAATTCACAGAGCAGAAAT 653

225 uValThrSerLeuGluAsnLeuArgAlaLysTyrTrpLeuAsnAlaGlyPheV 242

654 TGCCCAAGACATGGAAAAGTAACCGACTTCTACCAAGAACAGCGCTACT 703

242 AlArgPheGluIleLysAspAlaLysLeuAsnIleAsnGluAspLysAsn 258  
|||:||||| |||:||||| :|||:||||| :|||:||||| :|||:|||||

704 TCGATTTCCGCATCTCGATACCGACATCCAAACCCACGAGACAAAACC 753

259 ArgIlePheValGluIleSerLeuHisGluGlyGluGlnTyrArgPheGl 275  
:|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||

754 AAGCAGACCATCAAAATCAGCGTCCACGAAGCGGCGAGTTCCGTTGGG 803

275 yGlnThrGlnPheLeuGlyAsnLeuThr...TyrThrGlnAlaGluLeuG 291  
:|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||

804 CAAAGTCTCCATCGAAGCGCACCAACGAAGTCCCCAAAGCGGAATCGG 853

291 luAlaLeuLeuLysPheLysAlaGluGluGlyPheSerGlnAlaMetLeu 307  
|| ||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||

854 AAAAATCTCTACCATGAAGCCGGCAATGGTACGAACGCCACGACATG 903

308 GluGlnThrThrAsnAsnIleSerThrLysPheGlyAspAspGlyTyrTy 324  
:|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||

904 ACCGCGGTTTGGGTGAGATTTCAGAACCGCATGGCTCGGCGAGCTACGC 953

324 tTyrAlaGlnIleArgProValThrArgIleAsnAspGluSerArgThrV 341  
:|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||

954 ATACAGCAAAATACGCTACGCTGCGCAAGCGCGCAACCAACCAACCG 1003

341 AlAspValGluTyrTyrIleAspProValHisProValTyrValArgArg 357  
||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||

1004 TCGATTTGCTCTGCACATCGAACCGGCGCGGAAAAATCTAGCTCAAGAA 1053

358 IleAsnPheThrGlyAsnPheLysThrGlnAspGluValLeuArgArgGl 374  
:|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||

1054 ATPCCACATCAGCGGCAACAAACCAACCGCGACGAAGTCTGTCGCGCGCA 1103

374 uMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnLeuS 391  
:|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||

1104 ATTGGCCCAATGGAATCCCGCGCTTACGACACCTCCAAAGTCGACGCT 1153

391 eArgAlaArgLeuMetArgThrGlyPhePheLysHisValThrValAsp 407  
:|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||

1154 CCAAAGACGGCTGCGAGCTTTTGGGCTACTTCGACACGCTACAGTTTGAT 1203

408 ThrArgProValProAsnSerProAspGlnValAspValAsnPheValVa 424  
:|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||

1204 GCGGTCGCGCTTCGGGGCACACCGCAAGAAATCGCATTTGAACATGACCT 1253

424 lGluGlnProSerGlySerSerThrIleAlaAlaGlyTyrSerGlnS 441  
:|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||

1254 GACCCAAAGCTTCCACCGCTCGCTCGACTGAGCGCGGGCTGGGACAGG 1303

[illegible]

728 alPheAspThrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPhe 744  
 2150 TGTGGGACGGCAAAACCTACGACGACGACGAGTTCGCGGACC.....2193  
 745 LysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgPr 761  
 2194 .....GGCGGACGGGTTCAAACACATTATACGGCGCGGCAATACCCA 2234  
 761 oLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaT 778  
 2235 TAAATCCACCTTACCAACGAATTGCGCTATTCGCGCGCGCGGGTTA 2284  
 778 hrTrpTyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeu 794  
 2285 COTGGCTCTCGCTTTAGGCGCGGATGAAATTCAGTACGCTACCGCTG 2334  
 795 AsnLysLysGlnAsnAspClnThrAspThrValGlnPheGlnIleGlySe 811  
 2335 AAGAAAAACCGGAAGCAATCCACGCTTCCAAATTCACACTCGGCAC 2384  
 811 rValPhe 813  
 2385 GACGTTT 2391

seq\_name: /SIBS1/gcgdata/geneseq/geneseq-nembl/NA2001A.DAT:AAH42130

seq\_documentation\_block:

ID AAH42130 standard; DNA: 2394 BP.

AC AAH42130;

XX 17-SEP-2001 (first entry)

XX Nucleotide sequence of a Neisseria serogroup A protein.

XX Serogroup A protein; outer membrane protein; Neisserial infection;  
 XX vaccine; ss.

XX Neisseria meningitidis.

XX Key Location/Qualifiers

FT CDS 1..2394

FT /\*tag= a /product= "Neisseria serogroup A protein"

FT 1..63

FT /\*tag= b

FT mat\_peptide 64..2391

FT /\*tag= c

XX WO200152885-Al.

XX 26-JUL-2001.

XX 17-JAN-2001; 2001WO-IB00166.

XX 17-JAN-2000; 2000GB-0001067.

XX 09-MAR-2000; 2000GB-0005699.

XX (CHIR-) CHIRON SPA.

XX Pizza M, Rappuoli R, Giuliani M;

XX WPI; 2001-451895/48.

XX P-PSDB; AAB84746.

XX Composition for treating or preventing infection to, detecting, or for  
 XX raising antibodies against Neisserial bacteria, comprises an N.

XX meningitidis serogroup B outer membrane preparation and an immunogenic  
 XX component -

XX Disclosure; Page 70-71; 83pp; English.

XX The present sequence encodes a Neisseria serogroup A protein. The

CC protein is used to produce the compositions of the invention. The  
 CC specification describes a composition, comprising a Neisseria  
 CC meningitidis serogroup B outer membrane preparation and an immunogenic  
 CC component. The immunogenic component is protein disclosed in WO99/57280,  
 CC WO99/36544, WO99/24578, WO99/66791, WO97/28273, WO96/29412, WO95/03413,  
 CC WO99/31132, WO99/58683, WO99/55873, and/or N. meningitidis protein PorA,  
 CC ttpA, ttpB, pilC, opa, or omp85. The composition is used for making  
 CC a medicament for treating or preventing infection due to Neisserial  
 CC bacteria; a diagnostic reagent for detecting the presence of Neisserial  
 CC bacteria; or of antibodies raised against Neisserial bacteria; and/or  
 CC a reagent which can raise antibodies against Neisserial bacteria. It may  
 CC also be used as a vaccine.

XX  
 SQ Sequence 2394 BP; 655 A; 758 C; 586 G; 395 T; 0 other;

alignment\_scores:

Quality: 1111.00 Length: 819

Ratio: 2.035 Gaps: 14

Percent Similarity: 66.667 Percent Identity: 32.357

alignment\_block:

US-09-701-711-2 x AAH42130 ..

Align seg 1/1 to: AAH42130 from: 1 to: 2394

10 GlnValSerAlaMetThrMetAlaValMetMetValMetSerThrHisAl 26  
 4 AAACGTGAACACAGATTGCTTCCGCACTGATGGTCTGGGCATATCGCCITT 53  
 26 aglnAlaAlaAspPheMetAlaAsnAspIleThrIleThrGlyLeuGlnA 43  
 54 GGCACCTTCCGCACTTACCATTCCAAAGACATCGCTCGAAGGCTTGACG 103  
 43 rgValThrIleGluSerLeuGlnSerValLeuProPheArgLeuGlyGln 59  
 104 GTACCGAGCCGAGTACCTGATTCAACTACCTGCGCTCAAGTCGGCGAC 153  
 60 ValValSerGluAsnGlnLeuAlaAspGlyValLysAlaLeuTyrAlaTh 76  
 154 ACCTACACGACACACACGCGAGTGCCTCATCAAAAGCCTGTACGCCAC 203  
 76 rGlyAsnPheSerAspValGlnValTyrHisGlnGluGlyArgIleIleT 93  
 204 CGGTTCTTTGACGACGTACGCTCGCAAACTGCGGACGCGGACGCTCTGC 253  
 93 yrGlnValThrGluArgProLeuIleAlaGluIleAsnPheGluGlyAsn 109  
 254 TGACCGTTATCGACGCGCCACCACATCGCTCGCTCAACATCACCGCGCA 303  
 110 ArgLeuIleProLysGluGlyLeuGlnGluGlyLeuLysAsnAlaGlyLe 126  
 304 AAATGCTGCAAAACGACGCGCATTAAGAAAACCTCGAATCGTTCGGCT 353  
 126 uAlaValGlyGlnProLeuLysGlnAlaThrValGlnMetIleGluThrG 143  
 354 GCGCAGTCGCAATACCTTTAATCAGCGCACACTCAATCAGGCAGTCGCG 403  
 143 luLeuThrAsnGlnTyrIleSerGlnGlyTyrTyrAsnThrGluIleThr 159  
 404 GCCTGAAAGAGATACCTCGGCGCGCAAACTCAATATCCAAATCAGC 453  
 160 ValLysGlnThrMetLeuAspGlyAsnArgValLysLeuAspMetThrPh 176  
 454 CCCAAAGTAAACAACTCGCGCGCAACCGCTCGACATCGACATCAGCAT 503  
 176 eAlaGluGlyLysProAlaArgValValAspIleAsnIleGlyAsnG 193  
 504 TGACGAGGCGCAATCCGCAAAATCAGCGACATCGAATTTGAAGCGCAACC 553  
 193 lnHisPheSerAspAlaAspLeuIleAspValLeuAlaIleLysAspAsn 209  
 554 AAGTCTATCCGCGCAAACTGATGCGGCAGATGCTGTCGACCGAAGGC 603

```
210 LysIle...AsnProLeuSerLysAlaAspArgTyrThrGlnGluLysLeu 225
    |||   ::  |||   ::  |||   ::  |||   ::  |||   ::  |||   ::
604 GCAATTGGACATGGCTGCACAGCAACCAATTCACAGCAGAAATT 653
225 uValThrSerLeuGluAsnLeuArgAlaLysTyrLeuAsnAlaGlyPheV 242
    ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
654 TGCCCAAGACATGGAAAGATACCGACTTCACAGACAACACGGCTACT 703
242 aLArgPheGluIleLysAspAlaLysLeuAsnIleAsnGluAspLysAsn 258
    |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
704 TGAATTCGCGATCCTCGATACCGCATCCAAACCAACCAAGCAACCAACC 753
259 ArgIlePheValIleLysSerLeuHisGluGlyGluGlnTyrArgPheG 275
    ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
754 AAGCAGACCATCAAAATACCGTCCACGAAGCGGACGTTCCGTGGGG 803
275 yGlnThrGlnPheLeuGlyAsnLeuThr...TyrThrGlnAlaGluLeuG 291
    |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
804 CAAAGTCTCCATCGAAGCGGACACCAACGAAGTCCCAAGCCGAACCTGG 853
291 luAlaLeuLeuLysPheLysAlaGluGluGlyPheSerGlnAlaMetLeu 307
    ||  |||||  ::|||  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
854 AAAAAGTGTGACATGAAGCCCGCAATGTGTACGAACGCCAGCATG 903
308 GluGlnThrThrAsnAsnIleSerThrLysPheGlyAspGlyTyrTy 324
    ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
904 ACCGCGGTTTGGGTGAGATTCCAGAACCGCATGGGCTCGGCGGCTACGC 953
324 rTyrAlaGlnIleArgProValThrArgIleAsnAspGluSerArgThrV 341
    |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
954 ATACAGCGAAATCAGCGTACAGCGGTGCCAAACCGCCGCAACCAACCG 1003
341 aLAspValGluTyrTyrIleAspProValHisProValTyrValArgArg 357
    ||||  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
1004 TCGATTCTCTCTGCATCGACATCGAACCGCGCGGCAAAATCTACGTCA 1053
358 IleAsnPheThrGlyAsnPheLysThrGlnAspGluValLeuArgG 374
    |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
1054 ATCCACATCACCGGCAACCAACCAACCGCGACGAAAGTCTCGTGGCGGCA 1103
374 uMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnLeuS 391
    |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
1104 ATGGCGCAAAATGGAATCGCGCTTACACACCTCCCAAGCTGCAACGCT 1153
391 exArgAlaArgLeuMetArgThrGlyPhePheLysHisValThrValAsp 407
    ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
1154 CCNAGAGCGCTCGAGCTTTGGGCTACTTCGACACACGTACATTTGAT 1203
408 ThrArgProValProAsnSerProAspGlnValAspValAsnPheValVa 424
    ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
1204 GCGGTCCCGCTTGGCGGCACACCGCAAAATCGATTTCGAACATGAGCCT 1253
424 lGluGluGlnProSerGlySerSerThrIleAlaAlaGlyTyrSerGlnS 441
    |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
1254 GACCGCAAGTTCACCGGCTCGCTGACATTGAGCGCGGCTGGGTACAGG 1303
441 exGlyGlyValThrPheGlnPheAspValSerGlnAsnAsnPheMetGly 457
    ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
1304 ATACGCGCTGTGTCAGTCCGAGGCGTTTCCCAAGCAACCTGTTTCGGT 1353
458 ThrGlyLysHisValAsnAlaSerPheSerArgSerGluThrArgGluVa 474
    |||||  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
1354 ACGGGCAAGTCGCGCCCTCGCGGCTCAGCAAGCAACCAACCGCTCAA 1403
474 lTyrSerLeuGlyMetThrAsnProTyrPheThrValAsnGlyValSerG 491
    |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
1404 CGGCTCGCTGTGTTTACGACCGCTACTTCACGCGAGCGGCGGTACGCC 1453
491 lnsrLeuSerGlyTyrTyrArg.....LysThrLysTyrAspAsnLys 505
    ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
1454 TGGGTACGATGTTTACGGAAGAGCTTCGACCCCGCGCAACGATCGACC 1503
```

```
506 AsnIleSerAsnTyrValLeuAspSerTyrGlySerLeuSerTyrG 522
    ::|||::|||::|||::|||::|||::|||::|||::|||::|||::
1504 AGCATCAACAATATAAACACCACCGCAGCGCGCATCCGCATGAG 1553
522 yTyrProIleAspGluAsnGlnArgIleSerPheGlyLeuAsnAlaAsp 539
    |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
1554 CGTGCCCTGTTACCGAATACGCGCTGTAATTCGGTTTGGTGGCAGAAC 1603
539 snThrLysLeuHisGlyGlyArgPheMetGlyIleSerAsnValLysGln 555
    ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
1604 ACTGACCGCTCAAC.....ACCTACAACAAGCGCCCAAA 1638
556 LeuMetAlaAspGlyGlyLysIleGlnValAspAsnAsnGlyIleProAs 572
    |||||  |||   |||   |||   |||   |||   |||   |||   |||
1639 CACTATGCGGACTTTATCAAGAAATACGGCAAAACCGACGCG...ACAGA 1685
572 pPheLysHisAspTyrThrThrTyrAsnAlaIleLeuGlyTyrAsnTyrS 589
    |  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
1686 CGGCAGCTTCAAAGCGTGGCTGTACAAAGGTACCGTGGCTGGGGCGCA 1735
589 exSerLeuAspArgProValPheProThrGlnGlyMetSerHisSerVal 605
    ::::  |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||
1736 ACAAAACCCGACGCGGTTATGGCCACGCGGCTACCTGACGCGGCGTG 1785
606 AspLeuThrValGlyPhe...GlyAspLys...ThrHisGlnLysValVa 620
    ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
1786 AACGGCGAAATCGCCCTCGCGGACAGCAAACTGCAATACTACTCCGCCAC 1835
620 lTyrGlnGlyAsnIleTyrArgProPheIleLysLysSerValLeuArg. 636
    ::::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
1836 CCACAACCAACCTGTTCTTCCCTTAAAGCAAAACCTTACGCTGATGC 1885
637 .....GlyTyrAlaLysLeuGlyTyrGly.....AsnAsn 646
    ||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1886 TCGGCGCGAAGTCGCAATGCGGCG...GGCTACGCGCAGAACCAAGAA 1932
647 LeuProPheTyrGluAsnPheTyrAlaGlyGlyTyrGlySerValArgG 663
    ::|||::|||::|||::|||::|||::|||::|||::|||::|||::
1933 ATCCCTTCTTTGAAAACTTCTACGCGCGGCGCTGGGTTCGGTGGCGG 1982
663 yTyrAspGlnSerSerLeuGlyProArgSerGlnAlaTyrLeuThrAla 680
    |||||  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
1983 ATACGAAGCGGCGCGCTCGTCCGAAA..... 2010
680 rgArgGlyGlnGlnThrThrLeuGlyGluValVal.....GlyGlyAsn 694
    |||   |||   |||   |||   |||   |||   |||   |||   |||
2011 .....GTGTATGAGCAATACGGCGCAAAATCAGCTACGCGCGCAAC 2052
695 AlaLeuAlaThrPheGlySerGluLeuIleLeuProLeuProPheLysG 711
    |||   ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
2053 AAAAAAGCCAAAGCTCTCCGCGAGCTGCTTCCCGATGCCGCGCGCAA 2102
711 yAspTyrIleAspGlnValArgProValIlePheIleGluGlyGlnV 728
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
2103 AGAC...GCGCGCAGCTCGCGCTGAGCCTGTTGCCGACGCGGCGAGCG 2149
728 aPheAspThrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPhe 744
    |||   |||   |||   |||   |||   |||   |||   |||   |||
2150 TGTGGGACGGCAAAACCTACGACGCAACAGCAGTTCGCGGACCC..... 2193
745 LysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgPr 761
    ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
2194 .....GGCGGCGAGGTTCAAAACATTTACGGCGCGCGCAATACCCA 2234
761 oLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaT 778
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
2235 TAAATCCACCTTTACACAGAAATGCGCTATTCCCGCGCGCGCGGCTTA 2284
778 hrTyrThrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeu 794
    ||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
2285 CTGGCTCTCGCTTTAGGCGCGGATGAAATTCAGTACGCTACCCGCTG 2334
795 AsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGlnIleGlyse 811
```



854 AAAAAGTGTGACCATGAAGAAGCCCGCAAAATGGTAGAACGCCACGACGATG 903

308 GluInThrThrAsnAsnIleSerThrLysPheGlyAspAspGlyTyTr 324  
::: ::::::::::::::::::::| | | | |  
904 ACCGCGTTTGGTGAGATTTCAGAACCGCATGGCTCGGCAGGCTACGC 953

324 rTyrAlaGlnIleArgProValThrArgIleAsnAspGluSerArghTr 341  
| | | | | | | | | | | | | | |  
954 ATACAGCGAAAATTCAGCTACAGCGCTGCCAAACCGCAGAAACCAAACCG 1003

341 aAspValGluTyrrTyrIleAspProValHisProValTyrrValArgArg 357  
| | | | | | | | | | | | | | |  
1004 TCGATTTCGCTCGCACATCGAACCGCCGCGGAAATTCAGTCAACGAA 1053

358 IleAsnPheThrGlyAsnPheLysThrGlnAspGluValLeuArgArGGL 374  
| | | | | | | | | | | | | | |  
1054 ATCCACATCACCGGCAACAACAACCOCGACGAGTCTGTCGCCCGGA 1103

374 uMetArgGlnLeuGluGlyAlaLeuAlaSerAsnClnLysIleGlnLeus 391  
| | | | | | | | | | | | | | |  
1104 ATTGGCCCAAAATCGAATCCGGCGCTTACGACACCTCCAAGTCGAACGCT 1153

391 eArgAlaArgLeuMetArghrGlyPhePheLysHisValThrValAsp 407  
| | | | | | | | | | | | | | |  
1154 CCAAGAGCGGCTCGAGCTTTGGGCTACTTCGACACGTAACGTTTGT 1203

408 ThrArgProValProAsnSerProAspGlnValAspValAsnPheValVa 424  
| | | | | | | | | | | | | | |  
1204 GCGCTCCGCTTCGCGGCACACCCGACAAATTCGATTGAACATGACGCT 1253

424 lGluGluGlnProSerGlySerThrIleAlaAlaGlyTyrSerGlns 441  
| | | | | | | | | | | | | | |  
1254 GACCGAAGTTCACCGGCTCGCTCGACTTGAGCGCGGCTGGGTACAGG 1303

441 eRGIgLyValThrPheGlnPheAspValSerGlnAsnAsnPheMetGly 457  
| | | | | | | | | | | | | | |  
1304 ATACCGGCGCTGGTCATGTCGCGAGCGTTTTCCCAAGACAACCTGTCGGT 1353

458 ThrGlyLysHisValAsnAlaSerPheSerArgSerGluThrArgGluVa 474  
| | | | | | | | | | | | | | |  
1354 ACGGCGAGTCGGCGCCCTCGCGCTTCACGACGAAACCAACGACGCTCAA 1403

474 TyrSerLeuGlyMetThrAsnProTyrPheThrValAsnGlyLysSerG 491  
| | | | | | | | | | | | | | |  
1404 CGGCTCGCTGCTGTTTACGACCGCTACTTCACGCGACAGCGGGTACGCC 1453

491 InSerLeuSerGlyTyrrTyrArg.....LysThrLysTyrAspAsnLys 505  
| | | | | | | | | | | | | | |  
1454 TGGGTACGATGTTTACGGAAAGCCTTCGACCCGCGCAAGACATCGACC 1503

506 AsnIleSerAsnTyrrValLeuAspSerTyrrGlyGlySerLeuSerTyrgl 522  
| | | | | | | | | | | | | | |  
1504 AGCATCAACAATATAAAACCAACGCGACGCGCGCATCCCGCATGAG 1553

522 yTyrProIleAspGluAsnArgIleSerPheGlyLeuAsnAlaAspA 539  
| | | | | | | | | | | | | | |  
1554 CGTGCTGTTTACCGAATACGACCGCTGAATTCGGTTTGGTGGCAGAAC 1603

539 snThrLysLeuHisGlyGlyArgPheMetGlyIleSerAsnValLysGln 555  
| | | | | | | | | | | | | | |  
1604 ACCTGACCGCTCAAC.....ACCTACAACAAAGCGCCCCAAA 1638

556 LeuMetAlaAspGlyGlyLysIleGlnValAspAsnAsnGlyIleProAs 572  
| | | | | | | | | | | | | | |  
1639 CACTATGCGGACTTATCAAGAAATACGGCAAAACCCGACGCG...ACAGA 1685

572 pPheLysHisAspTyrrThrThrTyrrAsnAlaIleLeuGlyTripsAntyrs 589  
| | | | | | | | | | | | | | |  
1686 CGCGAGCTTCAAGGCTGGCTGTACAAAGGTACCGTCGGCTGGGGCGCA 1735

589 exSerLeuAspArgProValPheProthrGlnGlyMetSerHisSerVal 605  
| | | | | | | | | | | | | | |  
1736 ACAAAACCGACAGCGCTTATGGCGGACGCGGCTACCTGACGGGGCTG 1785

[illegible]

seq name: /SIDSL/qcdata/geneseq/geneseq-emb1/NA2000.DAT:AAA15155

seq documentation block:

seq\_documentation\_block.  
ID AAA15155 standard; DNA; 2505 BP.

XX  
XX

AAA15155;

XX  
XX

DT 21-AUG-20

XX

DE DNA encoded

XX

KW Outer mem

KW meningoco

XX

OS **Neisseria**

XX

•





```

1113 ATCCACATCCGCGCAACAACAAACCGGACGAAGTCGTGCGCGCGA 1162
374 uMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnLys 391
1163 ATTGCGCAATGGATCCGCGCTTACGACACTCTCAAGCTGCACGCT 1212
391 erArgAlaArgLeuMetArgThrGlyPhePheLysHisValThrValAsp 407
1213 CCAAGAGCGCGTCGAGCTTTGGGTACTTCGACAACTGACAGTTGAT 1262
408 ThrArgProValProAsnSerProAspGlnValAspValAsnPheValVa 424
1263 GCGTCCCGCTTCGCGGTACGCGCGCAAAAGTGTGATTTGAACATGAGCCT 1312
424 lGluGluGlnProSerGlySerThrIleAlaAlaGlyTyrSerGlns 441
1313 GACCGAACGTTCCACCGGCTCGCTGCAGCTTGACGGCGGTGGTTCAGG 1362
441 erGlyGlyValThrPheGlnPheAspValSerGlnAsnAsnPheMetGly 457
1363 ATACCGCTTGGTCAATGTCGCGCGGTATCGCAGGACAACTGTTGCGT 1412
458 ThrGlyIshisValAsnAlaSerPheSerArgSerGluThrArgGluVa 474
1413 ACGGCAAGTCGCGCGCTCGCGGCTCGCGAAGCAAAACCAACGCTCAA 1462
474 lTyrSerLeuGlyMetThrAsnProTyrPheThrValAsnGlyValSerG 491
1463 CGCTCGCTGCTGTTACCGACCGGTACCTTCAGCGACGCGGGTTCAGC. 1511
491 lnSerLeuSerGlyTyr.....TyrArgLysThr.....LysTyr 502
1512 .....CTGGCTACGATATTACGGAAGAGCCTTCGACCGCGCGCAA 1553
503 AspAsnLysAsnIleSerAsnTyrValLeuAspSerTyrGlyGlySerLe 519
1554 GCATCGACACGCGTCAACAAATATAAAACACCCCGCGCGCGCGCT 1603
519 uSerTyrGlyTyrProIleAspGluAsnGlnArgIleSerPheGlyLeuA 536
1604 AAGATGGGTATCCCGTTACCGAATACGACCGCGTCAATTCGGGCTGG 1653
536 snAlaAspAsnThrLysLeuHisGlyGlyArgPheMetGlyIleSerAsn 552
1654 CGCGGAACACCTGACCGTCAAC.....ACCTACAACAAA 1688
553 valLysGlnLeuMetAlaaspGlyGlyLysIleGlnValAspAsnAsnGl 569
1689 GCACCAAAAGCTATCCGACTTTATCAACAATACGCGCAAAACCGACGG 1738
569 ylleProAspPheLysHisAspTyrThrTyrAsnAlaIleLeuGlyT 586
1739 C...GCAGCGGAGCTTCAAGCGCTGTGTACAAAGGACACTGTCGGCT 1785
586 rAsnTyrSerSerLeuAspArgProValPheProThrGlnGlyMetSer 602
1786 GGGGCGCAACAAGACCGACGCGCTTATGGCGCGACGCGCGCTACCTG 1835
603 HisSerValAspLeuThrValGlyPhe...GlyAspLys...ThrHisGl 617
1836 ACCGGGTAATCCGAATCGCCCTGCGCGGACGAAACTGCAATACTA 1885
617 nLysValValTyrGlnGlyIleTyrArgProPheIleLysLysSerV 634
1886 CTCGGCCACCAACAACACCTGGTCTTCCCTTAAGCAAAACCTTCA 1935
634 alLeuArg.....GlyTyrAlaLysLeuGlyTyrGly... 644
1936 CGCTGATGCTCGCGCGGAAGTCGGCATTCGGGCG...GCTACGGCAGA 1982
645 ...AsnAsnLeuProPheTyrGluAsnPheTyrAlaGlyGlyTyrGlyse 660
.....

```

```

1983 ACCAAGAAATCCCTTCTTTGAAAACTTCTACGGCGCGGCTGGGTTCC 2032
660 rValArgGlyTyrAspGlnSerSerLeuGlyProArgSerGlnAlaTyrL 677
2033 GTTGGCGGCTACGAAGCGGACGCTCGGCCGAAA..... 2069
677 euThrAlaArgArgGlyGlnGlnThrThrLeuGlyGluValVal..... 691
2070 .....GTGTATGACGAATACGGCAAAATCAGCTAC 2102
692 GlyGlyAsnAlaLeuAlaThrPheGlySerGlnLeuIleLeuProLeuPr 708
2103 GCGGGCAACAAAGCAACGCTCGCGGAGCTCTTCCCGATGCC 2152
708 oPheLysGlyAspTyrIleAspGlnValArgProValIlePheIleGluG 725
2153 CGGTGCGAAAGAC...GCAGCACCGCTCGCGCTGAGCTGTTTCCCGCAGC 2199
725 lyGlyGlnValPheAspThrThrGlyMetAspLysGlnThrIleAspLeu 741
2200 CAGCAGCGTGTGG.....GACGGCAGAACCTATACCGCC 2234
742 ThrGlnPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAl 758
2235 GCCGAAACGTAACAAATCGTTTACTCGGAAACGCGATAAATC 2284
758 aAsnArgProLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyV 775
2285 C.....ACTTTACCAAGAAATGCGCTATTTCCCGCGCG 2319
775 alGlyAlaThrTyrThrProIleGlyProLeuSerIleSerTyrAla 791
2320 CGCGGTTACCTGCTCTGCTTGGCGCGATGAAATTCATCTACGCC 2369
792 LysProLeuAsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGl 808
2370 TACCGCTGAAGAAACCGGAAGACGAAATCCACGCTTCCAATTTCCA 2419
808 nileGlySerValPhe 813
2420 GCTCGGACGACGTTCC 2435

```

seq\_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:AAQ66199

seq\_documentation\_block:

ID AAQ66199 standard; DNA; 2984 BP.

XX AC AAQ66199;

XX DT 08-DEC-1994 (first entry)

XX DE H. influenzae b Eagen D15 sequence.

XX KW Vaccine; passive immunization; vector; antiserum; diagnosis; D15;

XX KW OMP; outer membrane protein; Hib; ss.

XX OS Haemophilus influenzae type b Eagen strain.

XX FH Key Location/Qualifiers

XX FT CDS 374..2767

XX XX /\*tag= a

XX PN WO9412641-A.

XX PD 09-JUN-1994.

XX PF 23-NOV-1993; 93WO-CA00501.

XX PR 23-NOV-1992; 92GB-0024584.

XX PA (CONN-) CONNAUGHT LAB LTD.

XX PI Chong P, Klein M, Loosmore S, Sia DYC, Thomas W;

PI Yang Y;  
 XX WPI; 1994-200269/24.  
 DR P-PSDB; AAR53755.  
 XX  
 PT Nucleic acid encoding D15 outer membrane protein - esp. of  
 PT Haemophilus influenzae, and related proteins, vectors, antisera  
 PT etc. useful in vaccines, for diagnosis and for passive  
 PT immunisation.  
 XX  
 PS Disclosure; Fig. 1B; 161pp; English.  
 XX  
 CC Outer membrane protein (OMP) D15 genes were isolated by screening  
 CC chromosomal libraries of H. influenzae type b (Hib) strains Ca,  
 CC Eagan and Mann A, and the non-typeable (NTHi) strains SB33 and PAK  
 CC 12085. Nucleotide sequences were determined for the D15 genes  
 CC (AAQ66198-202) and the corresponding aa sequences were derived  
 CC (AAR53754-58). D15 OMP can be produced easily and on a large scale,  
 CC free of other antigens and lipooligosaccharides, by recombinant DNA  
 CC methods using the isolated genes.  
 XX  
 XX Sequence 2984 BP; 960 A; 474 C; 619 G; 931 T; 0 other;

alignment\_scores:  
 Quality: 1052.50 Length: 822  
 Ratio: 1.956 Gaps: 20  
 Percent Similarity: 65.450 Percent Identity: 31.630

alignment\_block:  
 US-09-701-711-2 x AAQ66199 ..

Align seg 1/1 to: AAQ66199 from: 1 to: 2984

23 SerThrHisAlaGlnAlaAlaSpheMetAlaAsnAspIleThrIleth 39  
 413 ACAACGACTGCTTTGCCCGCACCCTTTGGGCAAGATATTTCGTGGA 462  
 39 rGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPheA 56  
 463 TGGTGTTCAAGTGACTTAGAACACAAATCCGAGCAAGTTTACCTGTTTC 512  
 56 rGlyLeuGlnValValSerGluAsnGlnLeuAlaAspGlyVallysAla 72  
 513 GTGCCCGGTGACGCTGTGATGCAATGATGTGGCTAAATATTTCGCTCT 562  
 73 LeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisGlnGluGl 89  
 563 TTATTCGTAAGTGGTTCGATTCGATGATGGAAGCG...CATCAAGAGG 609  
 89 yargile...ileTyrGlnValThrGluArgProLeuIleAlaGluIleA 105  
 610 CGATGTGCTGTTGTTAGGCTTGGCTAAATCGCATTCATTTCAGATGTTA 659  
 105 snPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluGlyLeu 121  
 660 AAATCAAGGTAACTCTGTATTTCCTCACTGAAGCACCTTAAACAAACTTA 709  
 122 LysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrValGl 138  
 710 GATGCTACCGGTTTAAAGTGGCGATGTTTAAATTCGAGAAAAATTA 759  
 138 nMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyrTyrA 155  
 760 TGAATTTGCCAAAGTGAAGAGACACTATCGAAGTGTAGTTCGCTATA 809  
 155 snThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgValLys 171  
 810 ACGCAACAGTTGAACCTATTGTCAATACGCTACCAATAATATCGCTGAA 859  
 172 LeuAspMetThrPheAlaGluGlyLysProAlaArgValValAspIleAs 188  
 860 ATTTTAATTCAAATCAATGAAGATGATGAAGCAAAATTCGCAATTAAC 909

188 nIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspValLeuA 205  
 910 TTTCAAGGGAACGAATCTGTTAGTAGCAGTACATTACAAGAACAAATGG 959  
 205 laIleLysAspAsnLysIleAsnProLeuSerLysAlaAspArgTyrThr 221  
 960 AATTACAACCTGATCTCTGGTGGAAATTA...TGGGAAATAAATTTGAA 1006  
 222 GlnGluLysLeuValThrSerLeuGluAsnLeuArgAlaLysTyrLeuAs 238  
 1007 GTGCGCAATTCGAAAGATTTCAGTCAATTCGFGATTATTATTATAA 1056  
 238 nAlaGlyPheValArgPheGluIleLysAspAlaLysLysIleAsnG 255  
 1057 TAATGGCTATGCCAAGACAAATTAATAACGGATGTTACAGTAAATG 1106  
 255 luAspLysAsnArgIlePheValGluIleSerLeuHisGluGlyGln 271  
 1107 ATGAAAAACAAAGTTAATGTAAACCATTTGATGTAATCAAGGTTTACAG 1156  
 272 TyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAl 288  
 1157 TATGACCTTCGTAGTGACGCAATTAAGTAATCTGGAGGTATGTCTGC 1206  
 288 aGluLeuGluAlaLeuLeuLys...PheLysAlaGluGluGlyPheSerG 304  
 1207 CGAGCTTGAACTTTACTTTTACAGCATTAATTAATGATCTTTCGCCG 1256  
 304 InAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAsp 320  
 1257 GTAGTGATATTCAGATGTAGAAATGCAATTAAGCAAAACTTGGAGAA 1306  
 321 AspGlyTyrTyrTyrAlaGlnIleArgProValThrArgIleAsnAspGl 337  
 1307 CCGGTTCGCTAGCGCAACGGAATTAATTCAGTACCTGATTTGATGATGC 1356  
 337 uSerArgThrValAspValGluTyrTyrIleAspProValHisProVal 354  
 1357 AAATAAAACATTAGCGATAACCTTTGTTGATGCTGGACGACGTTTAA 1406  
 354 yValArgArgIleAsnPheThrGlyAsnPheLysThrGlnAspGluVal 370  
 1407 CTTGTCGCCCACTTCGCTTTGAAGGAAATACCGTTTCTGCTGATGACAT 1456  
 371 LeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLy 387  
 1457 TTACGTCAAGAAATGCGCCCAACAGAGGAAGTGGTATAATTCACAAT 1506  
 387 sIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePheLysHisV 404  
 1507 ACTTGAGTTAGGAAAAATTCGCTTAGCTAGCTACAGGTTTCTTCGAA... 1552  
 404 alThrValAspThrArg.....ProValProAsnSerProAspGlnVal 418  
 1553 ..ACAGTCGAAACCGAATGTATCTCAATGCTAGTAGTAAATGATGAAGTG 1600  
 419 AspValAsnPheValValGluGlnProSerGlySerSerThrIleAl 435  
 1601 GATGCTGATATAAAGTCAAGACGTAACACGGGTAGTATCAACTTGG 1650  
 435 alaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSerG 452  
 1651 TATTGTTTACGGTACAGAGAGTGGTATTAGTTATCAAGCAAGTGTAAAC 1700  
 452 InAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468  
 1701 AAGATAATTTCTGGGACACGGGGCGGCAAGTATAGTATGCTGGTACGAA 1750  
 469 SerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheTh 485  
 1751 AATGATTATGTTACGAGTGTCAATTTGGGTTATACCGAGCCCTATTTTAC 1800

```
485 rValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLysT 502
1801 TAAAGATGGTGA.....AGTCTTGGGAAATGTTTCTTTGAAAACT 1844
502 yAspAsnLys.....AsnIleSerAsnTyrValLeuAspSerTyr 515
1845 ACGATAACTCTAAAGTGATACATCTCTACTATTAAGCGTAGGACTTAC 1894
516 GlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSe 532
1895 GGAAGTAATGTTACTTACTTGGTTCCTGCTGTAATGAAATAAATCCTTATTA 1944
532 rPheGlyLeu.....AsnAlaAspAsnThrLysLeuHisG 544
1945 TGTAGGATTAGGTACATACCTATAATAAATAGTAACCTTTGCTCTAGAAT 1994
544 lYcLYArgPheMetGlyIleSerAsnValLysGlnLeuMetaLaAspGly 560
1995 ATAACCGTAATTATATATATCAATCAATGAA..... 2026
561 GlyLysIleGlnValAspAsnAsnGlyIleProAspPheLysHisAspTy 577
2027 .....TTTAAAGGTAATGCAATAAACAAATGACTTTTGATT 2064
577 rThrThrTyrAsnAlaIleLeuGlyTrpAsnTyrSerSerLeuAspArgP 594
2065 TTCT.....TTGGTTGGAACTATACAGCCCTTAATAGAG 2099
594 roValPheProThrGlnGlyMetSerHisSerValAsp.....LeuThr 608
2100 GCTATTCCCACTAAAGGGTTAAAGCAAGTCTTGGTGGACGAGTTACT 2149
609 ValGlyPheGlyAspLysThrHisGlnLysValValTyrGlnGlyAsnI 625
2150 ATTCCAGGTTCTGATAACAATACTACAACTAAAGTCGACAGTGTACAGG 2199
625 eTyrArgProPhe.....IleLysLysSerValL 635
2200 TTCTACCCCTTAGACAGATACCTCTGGGTTGTATCTGCAAAAGCAT 2249
635 euArgGlyTyrAlaLysLeuGlyTyrGlyAsn....AsnLeuProPheTyr 650
2250 CTGCAGGATGCAAT...GGTTTGGAAACAAGCGTTTACCGTTCTAT 2296
651 GluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667
2297 CAAACTTATACAGCGGGTGGCATCGGTTTCATTACGTTGTTTGTATGG 2346
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
2347 TAGTATTGACCTAAGCA.....ATTATGCCGAATATGGTAATG 2387
684 lNThrThrLeuGly.....GluValValGlyGly 693
2388 GTAGTGTACTGTACTTTTAAAGAGATAAGTCTGATGTGATTGTTGTTG 2437
694 AsnAlaLeuAlaThrPheGlySerGluLeuIleLeuProLeuProPheLy 710
2438 AATGCAATCGCTACAGCTAGCCGACAGATTAAATTGCGCAACTCCCAPTT 2487
710 sGlyAspTrpIle...AspGlnValArgProValIlePheIleGluGlyG 726
2488 GAGCGATAGACCCAAATACGCTCGGAACCTCCTTATTTGTTGATCGCG 2537
726 lYcLInValPheAspThrThr...GlyMetAspLysGlnThrIleAspLeu 741
2538 CAAGTGTTTGGAATACTAAATGAAATCAGATAAAATGGATTAGAGGC 2587
742 ThrGlnPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAl 758
2588 GATGATTAAAA..... 2599
758 aAsnArgProLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyV 775
```

```
....|||
2600 .AGATTGCCTGATTATGGCAAAATCAAGCCGATATCGCGCTCTACAGGTG 2648
775 aGlyAlaThrTrpTyrThrProIleGlyProLeuSerIleSerTyrAla 791
2649 TCGGATTCCCAATGGCAATCTCTATTTGGGCATTTGTTCTCTATGCC 2698
792 LysProLeuAsnLysLysGlnAsnAspGlnThrAspThrValGlnPheG 808
2699 AAACCAATAAAAAATATGAAATGATGATGTCGACAGTTCCAATTTAG 2748
808 nIleGlySerValPhe 813
2749 TATGGAGGTTCTTTC 2764
seq_name: /SIDSI/gcdata/geneseq/geneseq-embI/NA1994.DAT:AAQ66198
seq_documentation_block:
ID AAQ66198 standard; DNA; 2949 BP.
XX
AC AAQ66198;
XX
DT 08-DEC-1994 (first entry)
XX
DE H. influenzae b Ca D15 sequence.
XX
KW Vaccine; passive immunization; vector; antiserum; diagnosis; D15;
KW OMP; outer membrane protein; Hib; ss.
XX
OS Haemophilus influenzae type b Ca strain.
XX
FH Key Location/Qualifiers
FT CDS 75..2468
FT FT /*tag= a
XX
PN W09412641-A.
XX
PD 09-JUN-1994.
XX
PF 23-NOV-1993; 93WO-CA00501.
XX
PR 23-NOV-1992; 93GB-0024584.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Chong P, Klein M, Loosmore S, Sia DYC, Thomas W;
PI Yang Y;
XX
DR WPI; 1994-200269/24.
DR P-PSDB; AAR53754.
XX
PT Nucleic acid encoding D15 outer membrane protein - esp. of
PT Haemophilus influenzae, and related proteins, vectors, antisera
PT etc. useful in vaccines, for diagnosis and for passive
immunisation.
XX
PS Disclosure; Fig. 1A; 161pp; English.
XX
CC Outer membrane protein (OMP) D15 genes were isolated by screening
CC chromosomal libraries of H. influenzae type b (Hib) strains Ca,
CC Eagan and Minn A, and the non-typeable (NTHi) strains SB33 and PAK
CC 12085. Nucleotide sequences were determined for the D15 genes
CC (AAQ66198-202) and the corresponding aa sequences were derived
CC (AAR53754-58). D15 OMP can be produced easily and on a large scale,
CC free of other antigens and lipooligosaccharides, by recombinant DNA
CC methods using the isolated genes.
XX
SQ Sequence 2949 BP; 983 A; 477 C; 612 G; 877 T; 0 other;
```

```
alignment_scores:
  Quality: 1051.50      Length: 822
  Ratio: 1.958          Gaps: 20
```

Percent Similarity: 65.328 Percent Identity: 31.630

alignment\_block:

US-09-701-711-2 x AAQ66198 ..

Align seg 1/1 to: AAQ66198 from: 1 to: 2949

```
23 SerThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIleThrIleTh 39
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
114 ACAAGGACTGTGTTGGCCGACCTTTTGGCAAGATATTTCGGTGA 163
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
39 rGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPheA 56
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
164 TGGTGTTCAAGGTGACTTTAGAACACAAATCCGAGCAAGTTTACCTGTC 213
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
56 rGLeuGlnValValSerGluAsnGlnLeuAlaAspGlyValLysAla 72
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
214 GTGGGTCACGGTGTGACTGACAAATGATGGCTAATATTGTCGGCTCT 263
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
73 LeuTyAlaThrGlyAsnPheSerAspValGlnValTyHisGlnGluG 89
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
264 TTATTCGTAAGTGTGATTCGATGATGTAAGCG...CATCAAGAAG 310
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
89 yArgIle...lleTyGlnValThrGluArgProLeuIleAlaGluIleA 105
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
311 CGATGTGCTTGTGTGATGCGTGTGGCTTAATTCGATCATTTTCAGATGTA 360
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
105 'snPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluGlyLeu 121
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
361 AAATCAAGGTAAGTCTGTTATCCCACTGAAGCATTAAACAAACATTA 410
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
122 LysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrValG 138
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
411 GATGCTAACGGTGTAAAGTTGGCGATGTTTAAATTCGAGAAATATAA 460
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
138 nMetIleGluThrGluLeuThrAsnGlnTyIleSerGlnGlyTyTyra 155
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
461 TGAATTTGCCAAAGTGTAAAGAGCACTATGCAAGTGTAGTGCCTATA 510
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
155 snThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgValLys 171
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
511 ACGCAACAGTTGAACCTATTGCTCAATAGCTACCAATAATTCGCGTGA 560
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
172 LeuAspMetThrPheAlaGluGlyLysProAlaArgValValAspIleAs 188
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
561 ATTTTAATTCAAATCAATGAAGATGATAAGCAAAATTTGGCATCATTAAC 610
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
188 nIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspValLeuA 205
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
611 TTTCAGGGGACCAAGTCTGTTAGTACGATACATTTCAAGAACAAATGG 660
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
205 laIleLysAspAsnLysIleAsnProLeuSerLysAlaAspArgTyThr 221
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
661 AATTACAACCTGATCTTGGTGGAAATTA...TGGGGAATAATAATTTGAA 707
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
222 GlnGluLysLeuValThrSerLeuGluAsnLeuArgAlaLysTyLeuAs 238
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
708 GGTCCGCAATTCGAGAAAGATTTCCAGTCATATTCGATATTATTATAA 757
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
238 nAlaGlyPheValArgPheGluIleLysAspAlaLysLeuAsnIleAsnG 255
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
758 TAATGGGTATGCCAAGACCAAAATTAATAAACGGATGTCAGCTAAATG 807
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
255 luAspLysAsnArgIlePheValGluIleSerLeuHisGluGlyGluGln 271
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
808 ATGAAAAACAAAAGTTAATGTAACCATTTGATGTAATGAAGTTTACAG 857
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
272 TyrArgPheGlyGlnThrGlnPheGlyAsnLeuThrTyThrGlnAl 288
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
858 TATGACCTTCGTAGTCAGCGCATTTATAGGTAATCTGGAGGTATGTCG 907
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
288 aGluLeuAlaLeuLeuLys...PheLysAlaGluGlyPheSerG 304
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
```

```
908 CGAGCTTGAACCTTTACTTTTCAGCATTACTTAAATGATACTTTCCGCC 957
304 lAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAsp 320
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
958 GTAGTGATATTGCAGATGTAGAAATGCAATTAAGCAAAACTTGGAGAA 1007
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
321 AspGlyTyTyTyTyAlaGlnIleArgProValThrArgIleAsnAspG 337
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1008 CGCGGTTACGGTAGCGCAACGGTAAATTCAGTACCTGATTTTGATGATGC 1057
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
337 uSerArgThrValAspValGluTyTyTyIleAspProValHisProValT 354
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1058 AAATAAAACATTTAGCATACCCCTTGTGTTGATGCTGCAGCAGCTTAA 1107
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
354 yValArgArgIleAsnPheThrGlyAsnPheLysThrGlnAspGluVal 370
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1108 CTGTCGCCAATTCGCTTTGAGGAAATACCGTTTCTGCTGATGACT 1157
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
371 LeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLy 387
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1158 TTACGTCAGGAAATGCGCCACAAGCAAGCACTTGGTATATTCACAATT 1207
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
387 sIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePheLysHisV 404
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1208 AGTTGAGTTAGGAAAAATTCGCTTAGATCGTACAGGTTTCTTCGAA... 1253
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
404 alThrValAspThrArg.....ProValProAsnSerProAspGlnVal 418
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1254 ..ACAGTCGAAACCGAATTCATCTATCAATGGTAGTAATGATGAAGTG 1301
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
419 AspValAsnPheValValGluGlnProSerGlySerSerThrIleAl 435
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1302 GATGTCGTATATAAGTCAAGAACGTAACACGGGTAGTATCACTTGG 1351
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
435 aAlaGlyTySerGlnSerGlyValThrPheGlnPheAspValSerG 452
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1352 TATTGCTTACGGTACAGAGAGTGGTATTAGTTATCAAGCAAGTGTAAAC 1401
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
452 lAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1402 AAGATAATTTCTGGGAACAGGGCGGCGAGTAAGTATAGCTGGTACGAA 1451
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
469 SerGluThrArgGluValTySerLeuGlyMetThrAsnProTyPheTh 485
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1452 AATGATTATGTAGCAGTGTCAATTTGGTTATACCGAGCCCTATTTTAC 1501
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
485 rValAsnGlyValSerGlnSerLeuSerGlyTyTyTyArgLysThrLysT 502
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1502 TAAAGATGGTGTA.....AGTCTTGGTGGAAATGTTTCTTTGAAAACT 1545
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
502 yAspAsnLys.....AsnIleSerAsnTyTyValLeuAspSerTyTy 515
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1546 ACGATAACTCTAAAAGTGATACATCTCTAACTATAAGCGTAGCATTAC 1595
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
516 GlyGlySerLeuSerTyTyTyTyProIleAspGluAsnGlnArgIleSe 532
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1596 GGAAGTAAATGTTACTTTAGGTTTCCCTGTAATGAAAAATAACTCTCTATTA 1645
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
532 rPheGlyLeu.....AsnAlaAspAsnThrLysLeuHisG 544
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1646 TGTAGGATTAGGTCATACCTATATAATAAATTAGTAACCTTCTCTAGAAT 1695
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
544 lyGlyArgPheMetGlyIleSerAsnValLysGlnLeuMetAlaAspGly 560
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1696 ATAACCGTAATTTATATATTCAATCAATGAAA..... 1727
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
561 GlyLysIleGlnValAspAsnAsnGlyIleProAspPheLysHisAspTy 577
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1728 .....TTTAAAGGTAATGGCAATTAAACAAATGACTTTTGATTT 1765
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
577 rThrThrTyTyTyTyAsnAlaIleLeuGlyTyTyTyTyTySerSerLeuAspArgp 594
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1766 TTCT.....TTTGGTGGAACTATAACAGCCTTAATAGAG 1800
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
```





```
...
2348 GTAGTGGTACTGGTACTTTTAAAGAGATAAGTCTGATGATGGTGGT 2397
694 AsnAlaLeuAlaThrPheGlySerGluLeuLeuLeuProLeuPheLeu 710
2398 AATCAATCCGCTACAGTAGCGCAGAGTTAATGTGCCAATCCATTTGT 2447
710 sGlyAspTrpIle...AspGlnValArgProValIlePheIleGluGly 726
2448 GAGCGATAAGAGCCAAATACGGTCCGAACTCTTATTGTTGATCGG 2497
726 lyGlnValPheAspThr...GlyMetAspLysGlnThrIleAspLeu 741
2498 CAAGTGTGGATACATAAATGGAATACAGATAAAATGGATTAGAGAGC 2547
742 ThrGlnPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLysAla 758
2548 GATGATTATATAA..... 2559
758 aAsnArgProLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGly 775
2560 .AGATTGCTGATATATGCAAAATCAAGCGTATTTCGCGCTCTACAGTG 2608
775 alGlyAlaThrTrpTyrThrProIleGlyProLeuSerIleSerTyrAla 791
2609 TCGGATTCCCAATGGCAATCTCTATTGGCCATTGGTATTCTCTATGCC 2658
792 LysProLeuAsnLysLysGlnAsnAspGlnThrAspThrValGlnPhe 808
2659 AAACCAATTAATAAATATGAAATGATGATGTCGAACAGTTCCAATTAG 2708
808 nIleGlySerValPhe 813
2709 TATTGGAGGTTCTTTC 2724
```

seq\_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:AAQ66202

seq\_documentation\_block:

```
ID AAQ66202 standard; DNA; 2989 BP.
AC AAQ66202;
XX
DT 08-DEC-1994 (first entry)
XX
DE H. influenzae PAK 12085 D15 sequence.
XX
KW Vaccine; passive immunization; vector; antiserum; diagnosis; D15;
KW OMP; outer membrane protein; NTHi; ss.
XX
OS Haemophilus influenzae PAK 12085.
XX
FH Key
FT CDS Location/Qualifiers
FT /*tag= a
XX
PN W09412641-A.
XX
PD 09-JUN-1994.
XX
PF 23-NOV-1993; 93WO-CA00501.
XX
PR 23-NOV-1992; 92GB-0024584.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Chong P, Klein M, Loosmore S, Sia DYC, Thomas W;
PI Yang Y;
XX
DR WPI: 1994-200269/24.
DR P-PSDB; AAR53758.
XX
PT Nucleic acid encoding D15 outer membrane protein - esp. of
PT Haemophilus influenzae, and related proteins, vectors, antisera
```

```
PT etc. useful in vaccines, for diagnosis and for passive
XX immunisation.
XX
PS Disclosure; Fig. 1E; 161pp; English.
XX
CC Outer membrane protein (OMP) D15 genes were isolated by screening
CC chromosomal libraries of H. influenzae type b (Hib) strains Ca,
CC Eagan and Minn A, and the non-typeable (NTHi) strains SR33 and PAK
CC 12085. Nucleotide sequences were determined for the D15 genes
CC (AAQ66198-202) and the corresponding aa sequences were derived
CC (AAR53754-58). D15 OMP can be produced easily and on a large scale,
CC free of other antigens and lipooligosaccharides, by recombinant DNA
CC methods using the isolated genes.
XX
SQ Sequence 2989 BP; 975 A; 468 C; 615 G; 931 T; 0 other;
```

alignment\_scores:

```
Quality: 1047.50 Length: 818
Ratio: 1.943 Gaps: 21
Percent Similarity: 65.892 Percent Identity: 31.663
```

alignment\_block:

US-09-701-711-2 x AAQ66202 ..

Align seg 1/1 to: AAQ66202 from: 1 to: 2989

```
23 SerThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIleThrIleTh 39
429 ACAACGACTGTGTTGGCGCACCTTTTGTGCAAAAGATATTCGCTGGA 478
39 rGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPheA 56
479 TGGTGTTCAGGTGACTTAGAACACAAATCCGAGCAAGTTTACCTGTC 528
56 rGlyLeuGlnValValSerGluAsnGlnLeuAlaAspGlyValLysAla 72
529 GTGCTGGTGGTGTGACTGACAATGATGGCTAATATTGTCGCTCT 578
73 LeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisGlnGlu 89
579 TTATTCGTAAGTGTGCTGATGATGTAAGCG...CATCAAGAAGG 625
89 yArgile...IleTyrGlnValThrGluArgProLeuIleAlaGluIleA 105
626 CGATGCTGCTGTTGTTAGCGTTGTGGTAAATCGATCATTCAGATGTA 675
105 snPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluGly 121
676 AATCAAGGTAAGTCTGTTATTCCTGCTGAAAGCACTTAAACAAACTTA 725
122 LysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrVal 138
726 GATGCTAACGGGTTTAAAGTTGGCGATGTTTAAATTCGAGAAAATTA 775
138 nMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyrTr 155
776 TGAATTTGCCAAAAGTGTAAAGAGACTATGCAAGTAGTGTAGTCTATA 825
155 snThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgVal 171
826 ACGCAACCGTTGAACCTATTGTCATACGCTGCCAAATAATCGTCTGAA 875
172 LeuAspMetThrPheAlaGluGlyLysProAlaArgValValAspIle 188
876 ATTTTAATTCAAATCAATGAAGATGATAAAGCAAAATTTGGCATTAAC 925
188 nIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspVal 205
926 TTTCAGGGGAACGAATCTGTGTAGCAGTACATTACAAAGCAAAATGG 975
205 laIleLysAspAsnLysIleAsnProLeuSerLysAlaAspArgTyrThr 221
```

976 AATTACAACCTGATTCTTGGTGGAAATTA...TGGGAAATAAATTTGAA 1022  
222 GlnGluLysLeuValThrSerLeuGluAsnLeuArgAlaLysTyrLeuAs 238  
1023 GUTGGCAATTCGAGAAAGATCTCGAGCAATTCGTGATTATTATTAA 1072  
238 nAlaGlyPheValArgPheGluIleLysAspAlaLysLeuAsnIleAsnG 255  
1073 TAATGGCTATGCCAAGACAAATCACTAAACGGATGTTACGTAATATG 1122  
255 luAspLysAsnArgIlePheValGluIleSerLeuHisGluGlyGln 271  
1123 ATGAAAAACAAGATTAATGTAACATTTAGTAAATGAAGTTTACAG 1172  
272 TyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAl 288  
1173 YATGACCTTCGTAGTCAGGCATTTATAGTAATCTGGGAGGTATGCTGC 1222  
288 aGluLeuGluAlaLeuLeuLys...PheLysAlaGluGluGlyPheSerG 304  
1223 CGAGCTTGAACCTTTACTTTACGATTTACATTTAAATGATACTTTCGCG 1272  
304 InAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAsp 320  
1273 GTAGTGATATTCAGATGTAGAAATTCGAATTAAGCAAAACTTGGGAA 1322  
321 AspGlyTyrTyrTyrAlaGlnIleArgProValThrArgIleAsnAspG 337  
1323 CGAGTTACGGTAAACACACACAGTAAATTCGTACCTGATTTGACAGTGC 1372  
337 uSerArgThrValAspValGluTyrTyrIleAspProValHisProValT 354  
1373 AAATAAACAATTAGCGATACCACTTTGTTGTTGATGCTGCAGCAGCTTAA 1422  
354 yValArgArgIleAsnPheThrGlyAsnPheLysThrGlnAspGluVal 370  
1423 CTGTTCCGCAACTTCGCTTTGAAGAAATACCGTTCTGCTGATAGTACT 1472  
371 LeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnly 387  
1473 TTACGTCAGAAATGCGACAAAGAAGAACTTGGTATATATCAAAAT 1522  
387 sIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePheLysHisV 404  
1523 AGTTGAGTTAGAAAATTCGCTTAGATCGTACAGGTTCTTCGAA... 1568  
404 alThrValAspThrArg...ProValProAsnSerProAspGlnVal 418  
1569 ..ACAGTTGAAAACCGAATTCATCTATCAATGGTAGCAATGATGAATG 1616  
419 AspValAsnPheValValGluGluGlnProSerGlySerSerThrIleAl 435  
1617 GATGTCGTATATAAAGTCAAGAAGCTAAACGGGTAGTATCAACTTTGG 1666  
435 aAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSerG 452  
1667 TATGTTGACGGTACAGAGTGGTATCATAGTTATCAACAAGTATTAAC 1716  
452 InAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468  
1717 AGATAAATTTCTTGGACACGGGGCGGAGTAAGTATATACCTGGTACGAA 1766  
469 SerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheTh 485  
1767 AATGATTATGTTGAGTGTCAATTTGGGTTATACCGAACCCCTATTTTAC 1816  
485 rValAsnGlyValSerGlnSerGlyTyrTyrArgLysThrLysT 502  
1817 TAAAGATGGTGTA....AGCTTGGTGGAAATATTTCTTTTGAAGAACT 1860  
502 yzAspAsnLys.....AsnIleSerAsnTyrValLeuAspSerTyr 515  
1861 ACGATAACTCTAAAGTGATACATCTCTAACTATAAGCGTAGCGACTTAT 1910

516 GlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSe 532  
1911 GGAAGTAAGTTACTTAGTTTCCTGTAAATGAAATAACTCTCTATTAA 1960  
532 rPheGlyLeu.....AsnAlaAspAsnThrLysLeuHisG 544  
1961 TGTAGATTAGGCCATACCTATAATAAATATAGTAACTTTGCTCTAGAAT 2010  
544 lyGlyArgPheMetGlyIleSerAsnValLysGlnLeuMetAlaAspGly 560  
2011 ATAACCGTAATTTATATATTCATCAATCAATGAAA..... 2042  
561 GlyLysIleGlnValAspAsnAsnGlyIleProAspPheLysHisAspTy 577  
2043 .....TTTAAAGGTAATGCATTAACAAACAAATGACTTTTGATT 2080  
577 rThrThrTyrAsnAlaIleLeuGlyTyrAsnTyrSerSerLeuAspArgp 594  
2081 TTCT.....TTTGGTTGGAACTATAACAGCCTTAATAGAG 2115  
594 roValPheProThrGlnGlyMetSerHisSerValAsp.....LeuThr 608  
2116 GCTATTTCCCACTAAAGGGTTAAAGCAAGCTTGTGGTGGACGAGTTACT 2165  
609 ValGlyPheGlyAspLysThrHisGlnLysValVal.....TyrGlnGl 623  
2166 ATTCCAGGTCTCTGATAACAAACTACTACAACTAAAGTGCAGATGTACAGG 2215  
623 y.....AsnIleTyrArgProPheIleLysLysSerValL 635  
2216 TTCTACCATTTAGACAGAGATCACCGCTGGGTGTGTTCTGCAAAAGCAT 2265  
635 euArgGlyTyrAlaLysLeuGlyTyrGlyAsn...AsnLeuProPheTyr 650  
2266 CTGCAGGATATGCAAT...GGTTTGGAAACAAGCGTTTACCCTTCTAT 2312  
651 GluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667  
2313 CAACTTATACAGCGGTGGCTGTTGTTTACGCGGTTTTCCTTATGG 2362  
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684  
2363 TAGTATTTGGCCTTAATGCA.....ATTATGCCGAACATGGTAAATG 2403  
684 InThr.....ThrLeuGlyGluValValIcIlyGlyAsnAlaLeuAla 697  
2404 GTACTTTTAAATAAGATAAGTTCGTGATGATGTTGTTGTAATGCAATCACA 2453  
698 ThrPheGlySerGluLeuLeuLeuProLeuProPheLysGlyAspTrpIl 714  
2454 ACTGCGAGTGCAGAACTTATGTACCAACTCCATTTGTGAGTGATATAAG 2503  
714 e...AspGlnValArgProValIlePheIleGluGlyGlyGlnValPhea 730  
2504 CCAAAATACAGCCGGAACCTCCCTATTTGTTGATGCGGCAAGTTTGGGA 2553  
730 spThrThr...GlyMetAspLysGlnThrIleAspLeuThrGlnPheLys 745  
2554 ATACTAAATGGAAATCAGATAAAATGGATTAGAGAGCAAGGCTTTGAAA 2603  
746 AspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLe 762  
2604 GAC.....TTACCTGA 2614  
762 uLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrT 779  
2615 TTATGGCAATCAAGCCGATTTCGCCCTCTACAGGTTGCGGATTCCAAT 2664  
779 rPtyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeuAsn 795  
2665 GGCAATCTCTATTGGACCATTTGTTATTTCTTATGCTAAACCAATAAA 2714





```

1319 CGAGGTTACGGTAAACACACAGATAAATCTGTACCTGATTTTGACGATGC 1368
      |||||      :   :   :   :   |||      :   :   :   |||
337 uSerArgThrValAspValGluTyrThrIleAspProValHisProValT 354
      :   :   :   :   :   :   :   :   :   :   :   :   :
1369 AAATAAAACATTAGCGATAACCTTTGTTGTGTGCTGGCAGCAGTTTAA 1418
      :   :   :   :   :   :   :   :   :   :   :   :   :
354 yrValArgIleAsnPheThrGlyAsnPheLysThrGlnAspGluVal 370
      |||||      :   :   :   :   |||||      :   :   :   :
1419 CTGTTCAACCACTTCGCTTGAAGGAAATACCGTTTCTGCTAGTAGTACT 1468
      :   :   :   :   :   :   :   :   :   :   :   :   :
371 LeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLy 387
      |||||      :   :   :   :   |||||      :   :   :   :
1469 TTACGTGCGAAGTATGCGCAACAAGGAAGAACTTGGTATAATTCACAATT 1518
      :   :   :   :   :   :   :   :   :   :   :   :   :
387 sIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePheLysHisV 404
      :   :   :   :   :   :   :   :   :   :   :   :   :
1519 AGTTGAGTTAGGAAATAATTCGCTTAGATCGTACAGGTTTCTTCGAA... 1564
      :   :   :   :   :   :   :   :   :   :   :   :   :
404 aIThrValAspThrArg.....ProValProAsnSerProAspGlnVal 418
      :   :   :   :   :   :   :   :   :   :   :   :   :
1565 ..ACAGTTGAAACCGCAATTGATCTCTATCAATGGTAGCAATGATGAAGTG 1612
      :   :   :   :   :   :   :   :   :   :   :   :   :
419 AspValAsnPheValValGluGluGlnProSerGlySerSerThrIleAl 435
      |||||      :   :   :   :   |||||      :   :   :   :
1613 GATGTCGTATATAAGTCAAGCAAGCAATACACGGGTAGTATCAACTTTGG 1662
      :   :   :   :   :   :   :   :   :   :   :   :   :
435 alaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSerG 452
      |||||      :   :   :   :   |||||      :   :   :   :
1663 TATTGGTTACGGTACAGAGAGTGGTATTAGTTATCAAGCAAGTGCAAC 1712
      :   :   :   :   :   :   :   :   :   :   :   :   :
452 InAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468
      |||||      :   :   :   :   |||||      :   :   :   :
1713 AAGATAAATTTCTTGGACACAGGGGGCGGCAAGTAAATAGTAGCTACGAAA 1762
      :   :   :   :   :   :   :   :   :   :   :   :   :
469 SerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheTh 485
      :   :   :   :   :   :   :   :   :   :   :   :   :
1763 AATGATTATGGTACGAGTGTCAATTTGGGTTATACCGAGCCCTATTATTAC 1812
      :   :   :   :   :   :   :   :   :   :   :   :   :
485 rValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLysT 502
      :   :   :   :   :   :   :   :   :   :   :   :   :
1813 TAAAGATGGTGA.....AGTCTTGGTGAATGTTTCTTTGAAAACT 1856
      :   :   :   :   :   :   :   :   :   :   :   :   :
502 yrAspAsnLys.....AsnIleSerAsnTyrValLeuAspSerTyr 515
      |||||      :   :   :   :   |||||      :   :   :   :
1857 ACGAAACTCTAAAGTAGTATACATCTCTAACTATAAGCGTACGACTTAT 1906
      :   :   :   :   :   :   :   :   :   :   :   :   :
516 GlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSe 532
      |||||      :   :   :   :   |||||      :   :   :   :
1907 GGAAGTAATGTTTACTTTAGGTTTCCTCTGTAATGAAATAACTCCTATTA 1956
      :   :   :   :   :   :   :   :   :   :   :   :   :
532 rPheGlyLeu.....AsnAlaAspAsnThrLysLeuHisG 544
      |||||      :   :   :   :   |||||      :   :   :   :
1957 TGTAGGATTAGCCCATACCTATATAATAAATATAGTAACCTTGTCTAGAA 2006
      :   :   :   :   :   :   :   :   :   :   :   :   :
544 lyClyArgPheMetGlyIleSerAsnValLysGlnLeuMetAlaAspGly 560
      :   :   :   :   :   :   :   :   :   :   :   :   :
2007 ATAACCGTAATTTATATATATCAATCAATGAAA..... 2038
      :   :   :   :   :   :   :   :   :   :   :   :   :
561 GlyLysIleGlnValAspAsnAsnGlyIleProAspPheLysHisAspTy 577
      :   :   :   :   :   :   :   :   :   :   :   :   :
2039 .....TTTAAAGGTAATGGCATTAAACAAATGACTTTGATTT 2076
      :   :   :   :   :   :   :   :   :   :   :   :   :
577 rThrThrTyrAsnAlaIleLeuGlyTrpAsnTyrSerSerLeuAspArgP 594
      :   :   :   :   :   :   :   :   :   :   :   :   :
2077 TTCT.....TTTGGTTGGAACATAACAGCCTTAATAGAG 2111
      :   :   :   :   :   :   :   :   :   :   :   :   :
594 roValPheProThrGlnGlyMetSerHisSerValAsp.....Leuthr 608
      |||||      :   :   :   :   |||||      :   :   :   :
2112 GCTATTTCACCAACTAAAGGGGTAAAGCAAGTCTTGGTGACGAGTTACA 2161
      :   :   :   :   :   :   :   :   :   :   :   :   :
609 ValGlyPheGlyAspLysThrHisGlnLysValValTyrGlnGlyAsnIl 625
      :   :   :   :   :   :   :   :   :   :   :   :   :

```

```

2162 ATTCCAGGTTCTGTATAACAATACTACAAACTAAGTGCAGATGTACAGGG 2211
      :   :   :   :   :   :   :   :   :   :   :   :   :
625 eTyrArgProPhe.....IleLysLysSerValL 635
      :   :   :   :   :   :   :   :   :   :   :   :   :
2212 TTTCATCCCATTAGACAGATACACCTCTGGGTTGTATCTGCAAAAGCAT 2261
      :   :   :   :   :   :   :   :   :   :   :   :   :
635 euArgGlyTyrAlaLysLeuGlyTyrGlyAsn....AsnLeuProPheTyr 650
      |||||      :   :   :   :   |||||      :   :   :   :
2262 CTGCAGATATGCAAAAT...GGTTTGGAAACAAGCGTTTACCGTTCTAT 2308
      :   :   :   :   :   :   :   :   :   :   :   :   :
651 GluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667
      :   :   :   :   :   :   :   :   :   :   :   :   :
2309 CAAACTATATACAGCGGTGTCATGTTCATTACCGGTTTCTCTTATGG 2358
      :   :   :   :   :   :   :   :   :   :   :   :   :
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
      :   :   :   :   :   :   :   :   :   :   :   :   :
2359 TAGCAATTGGCCCTAACGCA.....ATTATCAAGGTCAAA 2393
      :   :   :   :   :   :   :   :   :   :   :   :   :
684 InThrThrLeu.....GlyGluValValGlyGlyAsnAlaLeu 696
      :   :   :   :   :   :   :   :   :   :   :   :   :
2394 ATAATAAATTTAATAAGATAAGTTCTGTGTTGTTGTTGTAATGCAATC 2443
      :   :   :   :   :   :   :   :   :   :   :   :   :
697 AlaThrPheGlySerGlnLeuIleLeuProLeuProPheLysGlyAspTr 713
      |||||      :   :   :   :   |||||      :   :   :   :
2444 GCTACAGCTAGCGCAGAGTTAATTGCGCAACTCCATTGTGAGTGATAA 2493
      :   :   :   :   :   :   :   :   :   :   :   :   :
713 pile...AspGlnValArgProValIlePheIleGluGlyGlyGlnValP 729
      :   :   :   :   :   :   :   :   :   :   :   :   :
2494 GAGTCAAAATACAGTCCGAACCTCCCTATTGTTGATGCGGCAAGTGT 2543
      :   :   :   :   :   :   :   :   :   :   :   :   :
729 heAspThrThr...GlyMetAspLysGlnThrIleAspLeuThrGlnPhe 744
      :   :   :   :   :   :   :   :   :   :   :   :   :
2544 GGAATACTAAATGGAAATCAGATAAAATGGATTAGAGAGCAATGCTCTG 2593
      :   :   :   :   :   :   :   :   :   :   :   :   :
745 LysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgPr 761
      |||||      :   :   :   :   |||||      :   :   :   :
2594 AAAGAC.....TTACC 2604
      :   :   :   :   :   :   :   :   :   :   :   :   :
761 oLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlat 778
      :   :   :   :   :   :   :   :   :   :   :   :   :
2605 CGATTATGCAAAATCAAGCCGTACTCGCGCTCTACAGGTGTCGATTCC 2654
      :   :   :   :   :   :   :   :   :   :   :   :   :
778 hrTrpTyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeu 794
      |||||      :   :   :   :   |||||      :   :   :   :
2655 AATGGCAATCTCTCTAGTGACCAAGTGGTATTTCTTATGCTAAACCAAT 2704
      :   :   :   :   :   :   :   :   :   :   :   :   :
795 AsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGlnIleGlySe 811
      :   :   :   :   :   :   :   :   :   :   :   :   :
2705 AAAAAATATGAATGATGATGCGAACAGTCCCAATTTAGTATTGGGGG 2754
      :   :   :   :   :   :   :   :   :   :   :   :   :
811 rValPhe 813
      :   :   :   :   :   :   :   :   :   :   :   :   :
2755 TTCTTTC 2761
      :   :   :   :   :   :   :   :   :   :   :   :   :

```

seq\_name: /SIBS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AAF60966

seq\_documentation\_block:

ID AAF60966 standard; DNA; 1830 BP.

XX AAF60966;

XX AC

XX DT

XX 16-MAY-2001 (first entry)

XX P. putida KT2440-associated DNA ORF00612.

XX Transgenic plant; detection; probe; amplification; vaccine carrier;

XX microbial production strain; biological remediation; ds.

XX Pseudomonas putida.

XX DE19935088-A1.

XX PN

XX 01-FEB-2001.

PD



```

1122 CTGGCGTGCACCTACCTGATGACCACTCAAGACCCGCTTGTAGCCGC 1171
      |||||::: ||::: ||::: ||::: |||||
398 hrGlyPheLeuHisValThrValAspThrArgProValProAsnSer 414
      |||||::: |||||::: |||||::: |||||::: |||||
1172 TGGGCTTCTTCAAGGAAGTCAAGCTCGAGACCCGCGAGTGCCTGGCACT 1221
      |||||::: |||||::: |||||::: |||||::: |||||
415 ProAspGlnValAspValAsnPheValValGluGluGlnProSerGlyse 431
      |||||::: |||||::: |||||::: |||||::: |||||
1222 GACGACCAAGGTCGAGCTCACTACAGCTCGAAGAGCAGGCTCGGGCTC 1271
      |||||::: |||||::: |||||::: |||||::: |||||
431 rSerThrIleAlaLaGlyTyrSerGlnSerGlyValThrPheGlnP 448
      |||||::: |||||::: |||||::: |||||::: |||||
1272 CATCACCCGCGAGCTGCTTCCGACAGCGCGGCTGATCTCGGGTG 1321
      |||||::: |||||::: |||||::: |||||::: |||||
448 heAspValSerGlnAsnAsnPheMetGlyThrGlyLysHisValAsnAla 464
      |||||::: |||||::: |||||::: |||||::: |||||
1322 GTTCGATGACCCAGCAACTCTCTCGTACCGTGAACAGGTATCCATC 1371
      |||||::: |||||::: |||||::: |||||::: |||||
465 SerPheSerArgSerGluThrArgGluValTyrSerLeuGlyMetThrAs 481
      |||||::: |||||::: |||||::: |||||::: |||||
1372 GGCTGACCCGCTCGGAATACAGACCGGTTACAACCTCGGCTTCGTGA 1421
      |||||::: |||||::: |||||::: |||||::: |||||
481 nProTyrPheThrValAsnGlyValSerGlnSerLeuSerClyTyrTyrA 498
      |||||::: |||||::: |||||::: |||||::: |||||
1422 TCCCTACTTCACGCGGATGGCGTCTCCACAGATNGTT..... 1458
      |||||::: |||||::: |||||::: |||||::: |||||
498 rgLysThrLysTyrAspAsnLysAsnIleSerAsnTyrValLeuAspSer 514
      |||||::: |||||::: |||||::: |||||::: |||||
1459 .....GACGCCCATCTCTGAAACACAGGATGTCGACAAT 1494
      |||||::: |||||::: |||||::: |||||::: |||||
515 TyrGly 516.
      |||||
1495 CTTGGA 1500

```

seq\_name: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:AAS81841

seq\_documentation\_block:

ID AAS81841 standard; cDNA; 2057 BP.

XX AAS81841;

XX 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #17645.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX P-PSDB; ABG17654.

XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

XX Claim 1; SEQ ID No 17645; 103pp; English.

PS

XX

CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 2057 BP; 505 A; 495 C; 550 G; 507 T; 0 other;

alignment\_scores:

Quality: 901.00 Length: 648

Ratio: 2.110 Gaps: 9

Percent Similarity: 65.895 Percent Identity: 33.488

alignment\_block:

US-09-701-711-2 x AAS81841 ..

Align seg 1/1 to: AAS81841 from: 1 to: 2057

14 MetThrMetAlaValMetMetValMetSerThrHisAlaGlnAlaAAs 30

157 TTGCTCATAGCGTCGCTGCTGTTTGGCAGCGCCACCGTATACGGTGTGA 206

30 P...PheMetAlaAsnAspIleThrIleThrGlyLeuGlnArgValThrI 46

207 AGGTTCTGTAGTGAAGATATATTCATTTCGAAGCGCTTCAGCGTGTGCGG 256

46 leGluSerLeuGlnSerValLeuProPheArgLeuGlyGlnValValSer 62

257 TTGTTGCGGCCCTCTCTCAGTATGCGGTGCGCACAGGCGACACGTTAAT 306

63 GluAsnGlnLeuAlaAspGlyValLysAlaLeuTyrAlaThrGlyAsnPh 79

307 GATGAAGATATCAGTAATACCATTCGGGCTCTGTTTGTACCGCAACTT 356

79 eSerAspValGlnValTyrHisGlnGluGlyArgIleIleTyrGlnValT 96

357 TGAGGATGTTCCGCTCTTCTGATGTTGATACCCCTTCTGTGTCGTGTCG 406

96 hrGluArgProLeuIleAlaGluIleAsnPheGluGlyAsnArgLeuIle 112

407 AAGAACGTCGCAACATTCGCCAGCATTAATTCTCCGGTAACAATCGGTG 456

113 ProLysGluGlyLeuGlnGluGlyLeuLysAsnAlaGlyLeuAlaValG 129

457 AAAGATGACATGCTGTAAGCAAAACCTTCGAGGCTTATGTTGTGTCGGG 506

129 yGlnProLeuLysGlnAlaThrValGlnMetIleGluThrGluLeuThrA 146

507 CGAATCCCTCGATCGCACCCCAATTCGCCGATATCGAAGAGGTCGTGAAG 556

146 snGlnTyrIleSerGlnGlyTyrTyrAsnThrGluIleThrValLysGln 162

557 ACTTCTACTACAGCGTCGGTAAATATATAGCGCACGCTAAAGCTGTGCTG 606

163 ThrMetLeuAspGlyAsnArgValLysLeuAspMetThrPheAlaGluG 179





```

1171 GGGCCCTATTCCACCGTGGCAGGCTTACTATTCCGACCGCAGC.... 1215
683 nGlnThrThrLeuGlyGlnValValGlyGlyAsnAlaLeuAlaThrPheG 700
    ::::||||:|||||:|::|
1216 .....ACCGATGCTGGGGGCAATATCCTGTAAGGACGAG...AGCCAA 1299
700 lySerGluLeuLeuLeuProLeuProPheLysGlyAspTrpIleAspGln 716
    ::::||||:|||||:|||||
1253 GTCCGAGTACCTGTTCCTCCCTCCCTCGTGAAGGACGAG...AGCCAA 1299
717 valArgProValIlePheLeuGlyGlyGlnValPhe...AspThr.. 731
    ::::||||:|||||:|||||
1300 CTGCGCAGCTCGGTGCTGCTGCGGCGGACGACGTATGTCGCGACACCTG 1349
732 .....ThrGlyMetAspLysGlnThrIleAspLeuThrGlnPheL 745
    :|||:|:::|:::|:::|
1350 CTACCTGTCCACCCACCGAGGTCGCGCAGCGTCGACCTGGCG..... 1392
745 ysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgPro 761
1392 .....
762 LeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThr 778
    :|||:|:::|:::|:::|
1393 .....CAGATGGCTGTGCTGCTGGGGGTGGCGTGAC 1424
778 rTrpTyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeu 794
    :|||:|:::|:::|:::|
1425 CTGGTACAGCCCGATGGGCGGCTGAGTTTCAGCCCTGGGCGGCCATTA 1473

```

seq\_name: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AAI97964

seq\_documentation\_block:

ID AAI97964 standard; DNA; 6617 BP.

AC AAI97964;

DT 20-NOV-2001 (first entry)

DE Lawsonia intracellularis coding sequence SEQ ID NO: 1.

KW HtrA; PonA; HypC; YefW; ABCI; Omp100; Lawsonia intracellularis infection; vaccine; ds.

OS Lawsonia intracellularis.

PN JP20001169787-A.

PD 26-JUN-2001.

PF 20-OCT-2000; 2000JP-0320736.

PR 22-OCT-1999; 99US-0160922.

PA (PFIZ ) PFIZER PROD INC.

DR WPI; 2001-592540/67.

XX Lawsonia intracellularis polynucleotide and encoded protein, used to prevent Lawsonia intracellularis infection -

PS Claim 3; Page 37-39; 67pp; Japanese.

CC The present invention provides isolated polynucleotides encoding HtrA, PonA, HypC, YefW, ABCI or Omp100 protein of Lawsonia intracellularis. The sequences can be used in vaccines for the prevention of Lawsonia intracellularis infection. The present sequence is a coding sequence of the invention.

SQ Sequence 6617 BP; 2216 A; 1153 C; 1142 G; 2106 T; 0 other;

```

alignment_scores:
  Quality: 569.00      Length: 824
  Ratio: 1.284        Gaps: 27
  Percent Similarity: 53.762      Percent Identity: 23.908

alignment_block:
US-09-701-711-2 x AAI97964 ..
Align seg 1/1 to: AAI97964 from: 1 to: 6617

35 AspIleThrIleThrGlyLeuGlnArgValThrIleGluSerLeuGlnSe 51
   |||:|:::|:::|:::|:::|
4190 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4239
51 rValLeuProPheArgLeuGlyGlnValValSerGluAsnGlnLeuAla 68
   :|||:|:::|:::|:::|:::|
4240 ACGACATCATTAAATAAGGAGATCATCATCATCATCATCATCATCAT 4289
68 spGlyValLysAlaLeuTyrAlaThrGlyAsnPheSerAspValGlnVal 84
   :|||:|:::|:::|:::|:::|
4290 CAGAAATCAAAAAAATATGGAATTAGGATATTTAGTATGATGATGATG 4339
85 TyrHisGln.....GluGlyArg...IleIleTyrGlnValThrGl 97
   :|||:|:::|:::|:::|:::|
4340 AGATTATGAGAAACCGGGGAGGACGATTACTTGTATTACTGTACAGA 4389
97 uArgProLeuIleAlaGluIleAsnPheGluGlyAsnArgLeuIleProL 114
   :|||:|:::|:::|:::|:::|
4390 AAAGCCTAAATTAACAGATGTTGTTCAAGGCTCAAAAGCTGTAAAGTA 4439
114 ysGluGlyLeuGlnGluGlyLeuLysAsn...AlaGlyLeuAlaValGly 129
   :|||:|:::|:::|:::|:::|
4440 TCGATACATCTTCTGCTGCAATGAGTTCTAAAAAAGGATCAGTTATT 4489
130 GlnProLeuLysGlnAlaThrValGlnMetIleGluThrGluLeuThrAs 146
   :|||:|:::|:::|:::|:::|
4490 GATAGACTATTGTCCCAAGATATTCAA.....AAATATCCGA 4527
146 nGlnTyrIleSerGlnGlyTyrTyr.....AsnThrGluIleT 159
   :|||:|:::|:::|:::|:::|
4528 CCTCTATAGAAAGAGAGCTTACTATCTCGTGAAGTTAATTAATGAATA 4577
159 hrValLysGlnThrMetLeuAspLysAsnArgValLysLeuAspMetThr 175
   :|||:|:::|:::|:::|:::|
4578 AAGAGAAAGAA.....AATACTTCTTCGCAACACTATTGTTAACA 4618
176 PheAlaGluGlyLysProAlaArgValValAspIleAsnIleIleGlyAs 192
   :|||:|:::|:::|:::|:::|
4619 GTAAATGAGGGAAAAAATCTTATATTAAGATGTCGGAATTGAAGGACT 4668
192 nGlnHisPheSerAspAlaAspLeuIleAspValLeuAlaIleLysAsp. 208
   :|||:|:::|:::|:::|:::|
4669 TGAACAATAAAGCTTAAACTTTTAAAAAAGAGTTAGCATTAACAGAAC 4718
209 ..AsnLysIleAsnProLysSerLysAlaAspArgTyrThrGlnGluLys 224
   :|||:|:::|:::|:::|:::|
4719 GTAATTTTATCATGCTTACTTCTGGAACAGGTGTATTACGTGAAGATAT 4768
225 LeuValThrSerLeuGluAsnLeuArgAlaLysTyrLeuAsnAlaGlyPh 241
   :|||:|:::|:::|:::|:::|
4769 CTTGAAGCGTGACTCTATAGCAATCTCTGCCTATGCCATGAATCATGCTA 4818
241 eValArgPheGluIleLysAspAlaLysLeuAsnIleAsnGluAspLysA 258
   :|||:|:::|:::|:::|:::|
4819 TGTAGATATTCAAGTGTCTTCACTGAGTAACATTCATCAAGAA.....A 4862
258 snArgIlePheValGluIleSerLeuHisGluGlyGluGlnTyrArgPhe 274
   :|||:|:::|:::|:::|:::|
4863 AAGGAATTGTTATTACATTTAGAGTAAAGAGGTAAAGCGCTATAAATA 4912
275 GlyGlnThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAlaGluLeuGl 291
   :|||:|:::|:::|:::|:::|
4913 GGAAAAATAGACTTTAAAGGAGATCTTATTGAGACAAATGAACAACCTCT 4962

```









[illegible]

1938 ACCAGAAATAAGAGGTATTTGGGATAGGGATTAC...CATACGCCCTATCA 1984  
579 hrtyrasnAlaIleuGlyTrpAsnTyrSerSerLeuAspArgProVal 595  
1985 CCAGCTCTTCACCCCTGATGTGAGCTATGACAAACCGATGATTATTAC 2034  
596 PheProthrGlnGlyMetSerHisSerValAspLeuThrVal..... 609  
2035 TTCCTAGAAATGGGGTTATCTTTAGTTCCCTATGCGACGATGCTCGGCTT 2084  
610 .....GlyPheGlyAspLysThrH 616  
2085 GCCAAGCTCTGGCAGCTCAATCTTGAACGGGTATGGCGGAATGTCC 2134  
616 isGlnLysValValTyr.....GlnGlyAsnIleTyrArg 627  
2135 GTAACACCAAGTTTATGGTAAATTCGCCGCTTACCACCATTTGCCAAAA 2184  
628 ProPheIleLysLysSerValLeuArgGlyTyrAlaLysLeuGlyTyr.. 643  
2185 TATTTATTGATAGATTGATCGCTCGCTTTAAACACGAGGAGGTTATAT 2234  
644 .....GlyAsnAsnLeuProPheTyrGluAsnPheTyrA 655  
2235 CTTTAGGTATAACACCGCATGATTACTTGCCTTTAACTCCACCTTCTACA 2284  
655 laGlyGlyTyrGlySerValArgGlyTyrAspGlnSerSerLeuGlyPro 671  
2285 TGGGGGCGGTAACCCGGGTGAGAGGCTTTAGAACGGATCGGTATCTCCT 2334  
672 ArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnGlnThrThrLeuGln 688  
2335 AAAGATGAG.....TTTGG 2348  
688 yGluValValGlyGlyAsnAlaLeuAlaThrPheGlySerGlnLeu.... 703  
2349 CTTGTGCTTTGGAGCGATGGGATTTTACCGCTTCTACTGAATTGAGCT 2398  
704 .....IleLeuProLeuProPheLysGlyAspTrpIleAspGlnValArg 718  
2399 ATGGGGTGCTAAAGCGCGCT.....AAATGCGC 2427  
719 ProValIlePheIleGluGlyGlyGlnValPheAspThrThrGlyMetAs 735  
2428 TTAGCGGTGTTTTTGACTTTGGT..... 2451  
735 pLysGlnThrIleAspLeuThrGlnPheLysAspProGlnAlaThrAlaG 752  
2452 .....TTCTTAACTTTTAAACCCCA.....A 2473  
752 luGlnAsnAlaLysAlaAlaArgProLeuLeuThr..... 764  
2474 CTAGAGGGAGTTTTTCTATTAACGCTCTGTTACGACACGGAATTTAA 2523  
765 .....GlnAspLysGlnLeuArgTyrSe 772  
2524 GATTATGGCGTTATAGGCGCTGGGTTTGAAGACGCACTTGGAGGCGCTC 2573  
772 rAlaGlyValGlyAlaThrTrpTyrThrProIleGlyProLeuSerIleS 789  
2574 CACAGGCTTCAGATTGAATGGATTTCCGCCATGGGCGCTTTGGTGTGA 2623  
789 erTyr..... 790  
2624 TTTTCCCTATAGCTTTTTCACCAANTGGGGGATGGCAATGCCAAGAA 2673  
791 AlaLysProLeu.....AsnLysLysGlnAsnAspGlnThrAspThrVa 805  
2674 TGTAAAGGGCTATGCTTCAACCCCTAACATGGACGATTACACGCAACACT 2723  
805 lGlnPheGlnIleGlySerValPhe 813







230 uAsnLeuArgAlaLysTyrLeuAsnAla.....GlyPheValA 243  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 852 GTATATGCCTAGGGGTACTTAGACGCTCATATTTCTCGCCTTTTTTGA 901  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 243 iq.....PheGlulIleYsAspAlaLysLeuAsnIleAsnCluAsPls 257  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 902 AAACGGATTTCACCACCCATGACGCTAACGCTCCAT..... 936  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 258 AsnArgIlePheValclulIleSerLeuHisGluGlyGluGlnTyrArgPh 274  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 937 .....TATAAGGTCAAAGAGGGGATCCATACAGGAT 968  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 274 eGlyGlnThrGlnPhe...LeuGlyAsnLeuThrTyrThrGlnAlaGluL 290  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 969 TTCAGATATTTTAATAGAGATTGACAACCCGGTAGCTGCCCTTAAAAACCT 1018  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 290 euGluAlaLeuLeuLysPheLysAlaGluGluGlyPheSerGlnAlaMet 306  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 1019 TAGAAAACCGCTTAAGTTAAAGGAAGAAGTGCTTTTAATATTAGGCAT 1068  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 307 LeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAspAspGlyTY 323  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 1069 TTAAGAGCGGATCGCAATTTTAAAAACCGAAATCGCCGATAAGGGCTA 1118  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 323 rTyTrAlaGlnIleArgProValThrArgIleAsnAspGluSerArgT 340  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 1119 TCGGTTTGGGGTGGTGAAGCGACAGACTGGGTAAAGACGAAAAAACGGGC 1168  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 340 hrValAspValGluTyrTyrlleAspProValHisProValTyrValArg 356  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 1169 TTGTGAAGTCATTTATCGATTGAAGTGGCGATPATGGTCATATCAAT 1218  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 357 ArgIleAsnPheThrGlyAsnPhelysthrGlnAspGluValLeuArgAr 373  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 1219 GATGTCATATTTTAGGNACCGACGCGAGCGATAGATCATTTAGGAG 1268  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 373 gLUmetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnL 390  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 1269 GGAATGT.....TTACTAGGGCCTAAAGATAAATACAAC 1303  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 390 euSerArgAlaArg.....LeuMetArgThrGlyPhePhe 401  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 1304 TGACCAAATCTAGAAATTCGGAATAATCTTTGAGCGCTTAGGGTTTTTC 1353  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 402 LysHisValThrValAspThrArgProValProAsnSerProAspGlnVa 418  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 1354 TCATAAGTCACAGTTGAGAAAAAAGGTCAATAGCTCA.....TTGAT 1397  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 418 lAspValAsnPheValValGluGlnProSerGlySerSerThrIleA 435  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 1398 GGATTTGTTAGTGAGCGTAGAGAGGGCGCACCGGGCAGTTGCAATTCG 1447  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 435 laalagLyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValser 451  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 1448 GGTGGGCTATGGCTCTTATGGAGGGCTCATGCTTAATGGGAGCGTGAGC 1497  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 452 GlnAsnAsnPheMetGlyThrGlyLysHisValasn.....AlaSerPh 466  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 1498 GAAGGATCTTTTGGCACAGGGCAAGCATGAGCTTGATGTCAACAT 1547  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 466 eSerArgSerGluThrArgGlu..... 473  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 1548 TGCCACAGGGGGGTAGATCTTATCCGGGCATGCCAAAGGGGGGGC 1597  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 474 .....ValTyrSerLeuGlyMetThrAsnProTyTrPheThrValasn 487  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 1598 GTATGTTGCCGGGAATTTGAGCTTGACTAATCCAAGGATTTTGCACAGC 1647  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 488 GlyValSerGlnSerLeuSerGlyTyTrtyrargLysThrLysTyrAspAs 504  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 1648 TGGTATAGCTCTACGATCAATCTTTAT.....GCGGATTA 1682  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::

230 uAsnLeuArgAlaLysTyrLeuAsnAla.....GlyPheValA 243  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 852 GTATATGCCTAGGGGTACTTAGACGCTCATATTTCTCGCCTTTTTTGA 901  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 243 iq.....PheGluIleLysAspAlaLysLeuAsnIleAsnGluAspLys 257  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 902 AAACGGATTTCACCACCCATGACGCTAACGCTCCAT..... 936  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 258 AsnArgIlePheValGluIleSerLeuHisGluGlyGluGlnTyrArgPh 274  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 937 .....TATAAGGTCAAAGAGGGGATCCATACAGGAT 968  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 274 eGlyGlnThrGlnPhe...LeuGlyAsnLeuThrTyrThrGlnAlaGluL 290  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 969 TTCAGATATTTTAATAGAGATTGACAACCCGGTAGCTGCCCTTAAAAACCT 1018  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 290 euGluAlaLeuLeuLysPheLysAlaGluGluGlyPheSerGlnAlaMet 306  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 1019 TAGAAAACCGCTTAAGTTAAAGGAAGAAGTGCTTTTAATATTAGGCA 1068  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 307 LeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAspAspGlyTy 323  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 1069 TTAAGAGCGGATCGCAAATTTAAAAACCGAAATCGCCGATAAGGCTA 1118  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 323 rTyrTrAlaGlnIleArgProValThrArgIleAsnAspGluSerArgT 340  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 1119 TCGGTTTGGGTCGGTGAAGCGACAGACTGGGTAAAGACGAAAAAACGGGC 1168  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 340 hrValAspValGluTyrTyrIleAspProValHisProValTyrValArg 356  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 1169 TTGTGAAGTCATTTATCGATTGAAGTGGCGATPATGGTCATATCAAT 1218  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 357 ArgIleAsnPheThrGlyAsnPheLysThrGlnAspGluValLeuArgAr 373  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 1219 GATGTCATATTTACGGNACCGACGCGATAGATAGATCATTTAGAG 1268  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 373 gGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnL 390  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 1269 GGAATGT.....TTACTAGGCGCTAAAGATAAATACAAC 1303  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 390 euSerArgAlaArg.....LeuMetArgThrGlyPhePhe 401  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 1304 TGACCAACTAGAAATTCGGAATAATCTTTGAGCGCTTAGGGTTTTTC 1353  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 402 LysHisValThrValAspThrArgProValProAsnSerProAspGlnVa 418  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 1354 TCATAAGTCACAGTTGAGAAAAAAGGTCAATAGCTCA.....TTGAT 1397  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 418 lAspValAsnPheValGluGluGlnProSerGlySerSerThrIleA 435  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 1398 GGATTTGTTAGTGAGCGTAGAGAGGGCGCACCGGGCAGTTGCAATTCG 1447  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 435 laalagLyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSer 451  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 1448 GGTTCGGCTATGGCTCTTATGGAGGGCTCATGCTTAATGGGAGCGTGAGC 1497  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 452 GlnAsnAsnPheMetGlyThrGlyLysHisValasn.....AlaSerPh 466  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 1498 GAAGGATCTTTTGGCACAGGGCAAGCATGAGCTTGATGTCAACAT 1547  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 466 eSerArgSerGluThrArgGlu..... 473  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 1548 TGCCACAGGGGGGTAGATCTTATCCGGGCATGCCAAAAGGGCGGGC 1597  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 474 .....ValTyrSerLeuGlyMetThrAsnProTyrPheThrValasn 487  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 1598 GTATGTTTCCCGGAATTTGAGCTTGACTAATCCAAGGATTTTGCACAGC 1647  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 488 GlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLysTyrAspAs 504  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::  
 1648 TGGTATAGCTCTACGATCAATCTTTAT.....GCGGATTA 1682  
 :::: ||| |::: ||| |::: ||| |::: ||| |:::











```

675 CGTAGTGAAGTTCAAGTCAGCAGCAGCGCGCGTGGCCCTGGTCG 626
169 gVallYsLeuAspMetThrPheAlaGluGlyLysProAlaArgValVala 186
|||||:||||:||||:||||:||||:||||:||||:||||:||||:
625 CGTCAATCGCTTGGTCTCATCAATGAAGTGCACGAGAAATCGCGC 576
186 spLeAsnlelleGlyAsnGlnHisPheSerAspAlaAspLeuIleAsp 202
|||||:||||:||||:||||:||||:||||:||||:||||:||||:
575 CTATCAACTTCGTCGGTAAATCATGCTTATGGCGATAGCCCGCTGGCGGT 526
203 ValLeuAlaIleLysAspAsnLysIleAsnProLeuSer.....Ly 216
|||||:||||:||||:||||:||||:||||:||||:||||:||||:
525 GTTATCAGCACCAG.....AGTCCAAATCCTTGTGCTCTCTGACCCG 482
216 sAlaAspArgTyThrGlnGluLysLeuValThrSerLeuGluAsnLeuA 233
:||||:||||:||||:||||:||||:||||:||||:||||:
481 TAAGGATGCTATAAAGCAGCAGCAAGTTGAAGGCTGACGAGGAAGCTCTGC 432
233 rGAlaLysTyThrLeuAsnAlaGlyPheValArgPheGluIleLysAspAla 249
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
431 GTCAGTCTATTACAAATCATGTTATGCCGATTTCCGGATCAGCTCTTCG 382
250 LysLeuAsnIleAsnGluAspLysAsnArgIlePheValGluIleSerLe 266
:||||:||||:||||:||||:||||:||||:||||:||||:
381 GATCATCGCTGACGAGCAGCAGCAAGCAATACACCGCTCAACATCAGGT 332
266 uHisGluGlyGluGlnTyThrArgPheGlyGlnThrGlnPheLeuGlyAsnL 283
:||||:||||:||||:||||:||||:||||:||||:||||:
331 TGATGAAGTTCAGCGTTTACAAAGTTTCCGATATTAAATGTCGAAGCTCTG 282
283 eutHr...TyThrGlnAlaGluLeuGluAlaLeuLysPheLysAla 298
:||||:||||:||||:||||:||||:||||:||||:||||:
281 TTGAAGCGTGGCCCGCAGTGTGAGGCTGTGGTGCAGCATCGCGC 232
299 GluGluGlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsnIleSe 315
:||||:||||:||||:||||:||||:||||:||||:||||:
231 GGCGCGCTCTACAGCGCGCGCGCAATCAGAAATCGATGAAGCATCCA 182
315 rThrLysPheGlyAspAspGlyTyThrTyThrAlaGlnIleArgProValT 332
:||||:||||:||||:||||:||||:||||:||||:||||:
181 GCACGCGGTTTCCCGCAAGGTTATCCCTTTGCCCGCGGTCTCCGCGC 132
332 hrArgIleAsnAspGluSerArgThrValAspValGluTyThrIleAsp 348
:||||:||||:||||:||||:||||:||||:||||:||||:
131 GCAACCGCATATGGCAATGGCAGCATCGCGCTCACCTATATGTTGAC 82
349 ProValHisProValTyThrValArgIleAsnPheThrGlyAsnPheLy 365
:||||:||||:||||:||||:||||:||||:||||:||||:
81 CAGGGTGAGCGCGCTATGTCAGCGTATCGAAGTCAAGGCGCAATACAG 32
365 sThrGluAspGluValLeuArgArgGlu 374
:||||:||||:||||:||||:||||:||||:||||:||||:
31 GACACGGGATTACGTTATTCTCGCGAA 4
seq_name: /SDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAA61522
seq_documentation_block:
ID AAA61522 standard; DNA: 5740 BP.
XX
AC AAA61522;
XX
DT 23-OCT-2000 (first entry)
XX
DE A. vitis hypersensitive response elicitor-encoding DNA, SEQ ID NO:79.
XX
KW Hypersensitive response elicitor protein; HR: disease resistance;
KW insecticide; fungicide; antiviral; bactericide; growth enhancer;
KW stress resistance; transgenic plant; ds.
XX
OS Agrobacterium vitis.
XX
PN W0200028056-A2.
XX
```

```

4825 GGTAAAGCGTCTTACTCAACGGGTTATTTTCCGATGTGAAGATTTCGG 4874
      :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
86  isGlnGluGlyArgIleIleTyrGlnValThrGluArgProLeuIleAla 102
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
4875 TTTCCGGCAGCGCCCTGGTGTACGTCACGAGAACCAATTGGTCAAC 4924
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
103 GluIleAsnPheGluGlyAsnArgLeuIleProLeuGluGlyLeuGlnG1 119
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
4925 CAGGTGCTGTCAACGGCAACCGCAAGATCAAGATGACAAAGCTGACGAC 4974
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
119 uGlyLeuLysAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaT 136
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
4975 TGTGTGTACAGCGCAGCCACTGGT . . . . . CCTATAGCCAGGAAC 5015
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
136 hrValGlnMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGly 152
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
5016 TCATCAGCGCGATATCACCGCATCAAGCAGGCCTATCAGCCATTGGC 5065
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
153 TyrTyrAsnThrGluIleThrValLysGlnThrMetLeuAspGlyAsnAr 169
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
5066 CGTAGTGAAGTTGAAGTGACGACGACCGCGCGTGGCCCTGGTCG 5115
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
169 gValLysLeuAspMetThrPheAlaGluGlyLysProAlaArgValVala 186
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
5116 CGTCAATCTCGCTTTCGTCATCAATGAAGTGATCGCAGAAATCGGCG 5165
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
186 spIleAsnIleIleGlyAsnGlnHisPheSerAspAlaLeuIleAsp 202
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
5166 CTATCAACTTCGTCGTAATCATGCTTATGGCGATAGCCGCTTGGCGGT 5215
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
203 ValLeuAlaIleLysAspAsnLysIleAsnProLeuSer . . . . . Ly 216
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
5216 GTTATCAGCACCAAG . . . . . AAGTCCATCCTTCTGCTGACCCG 5259
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
216 salaAspArgTyrThrGlnGluLysLeuValThrSerLeuGluAsnLeuA 233
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
5260 TAAGGATGCTATAAGCAGCAGCAAGTTGAAGCTGACGAGAGCTCTGC 5309
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
233 rGAlaLysTyrLeuAsnAlaGlyPheValArgPheGluIleLysAspAla 249
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
5310 GTCAGTCTATTACATCATGTGTATGCGGATTTCCGGATACGCTTCG 5359
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
250 LysLeuAsnIleAsnGluAspLysAsnArgIlePheValIleSerLe 266
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
5360 GATGTCGCTGTAACGAGCAGACCAAGCAATACACCGTCAACATCAGGT 5409
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
266 uHisGluGlyGluGlnTyrArgPheGlyGlnThrGlnPheLeuGlyAsnL 283
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
5410 TGATGAAGGTCAGCGTTACAAAGTTTCCGATATTATATGTCGAAAGCTCG 5459
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
283 euThr . . . TyrThrGlnAlaGluLeuAlaLeuLeuLysPheLysAla 298
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
5460 TTGAAGCGGTGACCGCTGATGTTGAAGGTTGTTGTCGACGACATCGCCG 5509
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
299 GluGluGlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsnIleSe 315
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
5510 GCGGCGCTCTACAGCGCCGCGAATCCAGAAATCGATGGAACCATCCA 5559
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
315 rThrLysPheGlyAspGlyTyrTyrTyrAlaGlnIleArgProValT 332
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
5560 GCAGCGCTTTCGCCAAGGTTATCCCTTGGCCGCGTTCGTCGCGCG 5609
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
332 hrArgIleAsnAspGluSerArgThrValAspValGluTyrIleAsp 348
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
5610 GCAACCGCGATATGGCAATGCGACCATCGGCGTACCTATATGTTGAC 5659
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
349 ProValHisProValTyrValArgIleAsnPheThrGlyAsnPheLy 365
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
5660 CAGGTGACGCGCCCTATGTGACGCTATCGAAGTCAAGGCAATACAAG 5709
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
365 sThrGlnAspGluValLeuArgArgGlu 374
      : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

5710 GACACGCGATTACGTTATTTCGTCGCAA 5737
seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAC81914
seq_documentation_block:
ID AAC81914 standard; DNA; 273254 BP.
XX AAC81914;
XX 27-FEB-2001 (first entry)
XX Chlamydia pneumoniae genome DNA.
KW Genome; diagnosis; vaccine; ds.
XX Chlamydia pneumoniae.
XX WO200027994-A2.
XX 18-MAY-2000.
XX 12-NOV-1999; 99WO-US26923.
XX 12-NOV-1999; 98US-0108279.
XX 08-APR-1999; 99US-0128606.
XX (REGC ) UNIV CALIFORNIA.
XX Stephens R, Mitchell W, Kalman S, Davis R;
XX WPI; 2000-376516/32.
XX Isolated nucleic acid for use in diagnostic and analytical methods
XX encodes genomic sequence of Chlamydia pneumoniae -
PS Claim 2; Page 128-320; 320pp; English.
XX This invention describes a novel nucleic acid (N1) encoding a Chlamydia
XX pneumoniae protein (p1), given in the specification. The isolated nucleic
XX acid is useful for diagnostic and analytical methods, such as,
XX hybridization-based assays or amplification-based assays. The protein may
XX be used for diagnostic purposes, for their enzymatic or structural
XX activity, or as a vaccine. The invention also describes (1) a probe
XX comprising a hybridizing fragment of N1; (2) an isolated nucleic acid
XX (N2) that hybridizes under stringent conditions to N1; (3) an expression
XX cassette comprising N1 under the transcriptional regulation of a
XX transcriptional initiation region functional in an expression host, and a
XX transcriptional termination region; (4) a cell comprising an expression
XX cassette of (3) as part of an extrachromosomal element or integrated into
XX the genome of a host cell as a result of induction of the expression
XX cassette into the host cell, and the cellular progeny of the host cell;
XX (5) a method for producing a p1 comprising growing a cell of (4) where
XX the protein is expressed and isolating the protein free of other
XX proteins; (6) a purified polypeptide composition comprising at least 50
XX weight % of p1; and (7) a monoclonal antibody binding specifically to the
XX peptide of (6).
XX Sequence 273254 BP; 76423 A; 51054 C; 61965 G; 83812 T; 0 other;

```

```

alignment_scores:
Quality: 395.00 Length: 881
Ratio: 0.855 Gaps: 36
Percent Similarity: 52.440 Percent Identity: 22.020
alignment_block:
US-09-701-711-2 x AAC81914/rev ..
Align seg 1/1 to reverse of: AAC81914 from: 1 to: 273254
1 MetArgAsnSerTyrPheLysGlyPheGlnValSerAlaMetThrMetal 17
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
268517 ATCGAAATAAAGTTATCTTGGCAATATCTATTCTAGCGTTAATCCAAC 268468

```

[illegible]

266766 AGTGTC.....A 266759  
 585 lYrPAsnTyrSerSerLeuAspArgProValPheProThrGlnGlyMet 601  
 266758 ACTTGAATTACGATCTCTGTAGATAGTCCTAGAACCTCCAACACAGGGATT 266709  
 602 SerHisSerValAspLeuThrVal...GlyPheGlyAspLysThrHis... 616  
 266708 CCGGGGGGGTACCTTTTGGAGTTCCTGTTGGGAGAACTTATCATTT 266659  
 617 .GlnLysValValTyrGlnGlyAsnIleTyrArgProPheIleLysLys 633  
 266658 TCAAAACTCTCTTAAACAGCTCTATCTATAGAAAACCTTACGGCGTAAAG 266609  
 633 erValLeu.....ArgGlyTyrAlaLysLeu.....GlyTyrGlyAsn 645  
 266608 GTATTTTGAATCAAGGGAAGCTCAATTTATTAACCTCATAGCAAT 266559  
 646 .....AsnLeuProPheTyrGluAsnPheTyrAlaGlyGlyTy 658  
 266558 ACTACAGCTGAAGGAGTTCCTCTCAGTGAGCGCTTCTCTAGTGGAGA 266509  
 658 rGlySerValArgGlyTyrAspGlnSerSerLeuGlyProArgSerGlnA 675  
 266508 GACTACAGTTCGGGATATAAATCTTTATTATCGGTCCAAATACTCTG 266459  
 675 laTyrLeuThrAlaArgArgGlyGlnGlnThrThrLeuGlyGluValVal 691  
 266458 CT.....ACAGACCTCAG 266445  
 692 GlyGlyAsnAlaLeuAlaThrPheGlySerGluLeuLeuProLeuPr 708  
 266444 GGAGGACTCTCTCGCTCTCTTATTCAGAAGAGTTCAATACCCTCTC... 266397  
 708 oPheLysGlyAspTrpIleAspGln.....ValArgProValIlePheI 723  
 266396 .....ATCAGACACACCTAAATATTAGTCGCCCTTTGTATTCT 266363  
 723 leGluGlyGlyGlnValPheAspThrThrGlyMetAspLysGlnThrIle 739  
 266362 TAGACTCAGTTTGTCT.....GGTTTACAGAGTATAAGATT 266325  
 740 AspLeuThrGlnPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLy 756  
 266324 TCGTTA..... 266319  
 756 sAlaAlaAsnArgProLeuLeuThrGlnAspLysGlnLeuArgTyrSera 773  
 266318 .....AAAGATCTAGCTAGTAGTG 266300  
 773 laGlyValGlyAlaThrTrpTyrThrProIleGly...ProLeuSerIle 788  
 266299 CTGGATTGGTCTGGCTTCGATGTAATGAATAATGTTCTGTTATGTTA 266250  
 789 SerTyrAlaLysPro.....LeuAsnLysLysGlnAs 799  
 266249 GGATTTGGTGGCCCTCCGTCACACCGAGACTTTGATGAGAAATAAT 266200  
 799 nAspGlnThrAspThrValGlnPheGlnIleGlySerValPhe 813  
 266199 TGATGTATCTCAGCGATTCTCTTTGCTTTAGGGGGCATGTC 266157  
 seq\_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:AAAX91990  
 seq\_documentation\_block:  
 ID AAX91990 standard; DNA; 1230025 BP.  
 XX  
 AC  
 XX  
 XX  
 DT 13-SEP-1999 (first entry)  
 DE Nucleotide sequence of the complete genome of Chlamydia pneumoniae.

361 AlaLeuAlaSerAsnGlnLysIleGlnLeuSerArgAlaAArgLeuMetAr 397  
349245 GATACATTCAATCGCTTAAAGCTAGAGTACTGACCAACGTTTAAAGAA 349294  
397 gThrglyPhePheLysHisValThrValAspThr.....A 409  
349295 TACAGGCTACTTCCAAAGCGTTAGTGTCTATACAGTTCGTCTCAACTTG 349344  
409 rGProValProAsnSerProAspGlnValAspValAsnPheValValGlu 425  
349345 ATCTATATGGCAATCGCGATCAATACCGAGATATTTTGTAGAGTCAAA 349394  
426 GluGlnProSerGlySerSerThrIleAlaAlaGlyTyrSerGlnSerG1 442  
349395 GAACACAAACAGAGAACTTAGGCTTATCTTGAAGATTAGTTCCTTGA 349444  
442 yGlyValThrPheGlnPheAspValSerGlnAsnAsnPhe..... 455  
349445 CAATCTTTTGGAGGAATTTGAATATCTGAAAGTAATTTTGTATCTATTG 349494  
456 .....MetGlyThrGly 459  
349495 GAGCTAGAAATATATTTTCTAAAGTTTTCGTCTCTAAGAGCGCGTGA 349544  
460 LysHis.....ValAsnAlaSerPheSerArgSerGluThrArgGluVa 474  
349545 GAACATCTATTTTAAAGACCAACTTCGGGACAAAGTCACAGAC..... 349589  
474 lTyrSerLeuGlyMetThrAsnProTyrPheThr.....Vala 487  
349590 .TATACTTTGAAGTCGACCAACCTCATTTTCTAAACACTCTCTTGATTT 349638  
487 snGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLysTyrAsp 503  
349639 TAGGAATTTGAATTAGAT.....AAATCAATT 349664  
504 AsnLysAsnIleSer...AsnTyrValLeuAspSertYrGlyGlySerLe 519  
349665 AACAGAGCATTTATCTAAAGATTATCTGTCTCAAAACCTATGCGGGAACGT 349714  
519 uSerTyrGlyTyrProIleAspGluAsnGlnArgIleSerPheGlyLeuA 536  
349715 CAGCACACGTATATCTTGAACAGAACCTGAAATACCGTCTATTATTATC 349764  
536 snAlaAspAsnThrLysLeuHisGlyGlyArg...PheMetGlyIleSer 551  
349765 GAGGAAGTCAAAACGAGTTTACATGAAAAACCTTAAGTTCTCTCAGGGCA 349814  
552 AsnValLysGlnLeuMetAlaAspGlyGlyLysIleGlnValAspAsnAs 568  
349815 AATATA.....GACAGCAATAAAGGATTGTCTCTCGCTGTC 349849  
568 nGlyIleProAspPheLysHisAspTyrThrTyrAsnAlaIleLeuG 585  
349850 AGGTGTC.....A 349857  
585 lyTrpAsnTyrSerSerLeuAspArgProValPheProThrGlnGlyMet 601  
349858 ACTTGAATTACGATTCTGTAGATAGTCTAGAACTTCAAACTACAGGGATT 349907  
602 SerHisSerValAspLeuThrVal...GlyPheGlyAspLysThrHis... 616  
349908 CGCGGGGGTGACTTTTGGAGTTTCTGGTTGGAGGAAGTATATCAATT 349957  
617 .GlnLysValValTyrGlnGlyAsnIleTyrArgProPheIleLysLys 633  
349958 TACAAAACTCTCTTAAACAGCTCTATCTATAGAAAACTTACGCGTAAAG 350007  
633 erValLeu.....ArgGlyTyrAlaLysLeu.....GlyTyrGlyAsn 645  
350008 GTATTTTGAATAATCAAAAGGGAGCTCAATTTATTAAACCCCTATAGCAAT 350057  
646 .....AsnLeuProPheTyrGluAsnPheTyrAlaGlyGly 658







SQ Sequence 2550 BP; 814 A; 487 C; 478 G; 771 T; 0 other;

alignment\_scores:

Argument_Scores:		
Quality:	393.00	Length: 881
Ratio:	0.852	Gaps: 36
Percent Similarity:	52.327	Percent Identity: 22.020

alignment\_block:

US-09-701-711-2 x AAZ61502

Align seg 1/1 to: AA261502 from: 1 to: 2550

```

1  MetArgAsnSerTyrPheLysGlyPheGlnValSerAlaMetThrMetAl 17
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
101 ATGCCGAAAAAGATGTTCTCTGCAAAATATCTATTCTAGCGTTAATCAAAAC 150
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
17  aValMetMetValMetSerThrHis...AlaGlnAlaAlaAspPheMetAl 33
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
151 CCCCTTAACTTTTATTCTTCTACTGAAAAGGTTTAAAGAAGCGCATGTGGTGG 200

```

```

33 laAsnAspIleThr...IleThrGlyLeuGlnArgValThrIleGluSer 48
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
201 TAGACTCTATCAACATCATACGGAGGAGAAAATGCTCAATAAACAAT 250
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

49 LeuGlnSerValLeuProPheArgLeuGlyGlnValValSerGluAsnG1 65
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
251 CCCATTACCCAAATTAAGACACCAAGATGGGGCTCTTTTCTCAAAATAGA 300
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

65 nLeuAlaAspGlyValLysAlaLeuTyAlaThrGlyAsnPheSerAsp 82
301 TTTTGTGAAGACTTGAAGATTCTAGCTAAAGAAATACGACTCTGTTGAGC 350
82 alGlnValTyHisGlnGluGlyArg....IleIleTyGlnValThr 96
351 CTTAAAGTAGAATTTTCTGAAGGAGAAACTAACATAGCCCTTCACCTAATA 400

```

97 GluArgProLeuIleAlaGluIleAsnPhesGluGlYAsnAArgLeuIlePr 113  
 : : : : :  
 401 GCTAAACCCCAATTCGAAAATTTCATATCTCAGGAATCAAGTCGTCC 450  
 : : : : :  
 113 oLYSGlUGlUcInclucGlYLeuLySAsnAlaGlYLeuAlaValGlYG 130  
 : : : : :  
 451 TGAACATAAAATT .....CTTAAACCCTACAATTTTACCGTAATG 491  
 : : : : :

```

130 lnProLeuLysGlnAlaThrValGlnMetileGluThrGluLeuThrAsn 146
      : : : : : : : : : : : : : : : : : : : : : : : :
492 ATCTCTTGAACGAGAAAAATTTCTTAAGGCTCTTCATGATCTAAGAAGC 541
      : : : : : : : : : : : : : : : : : : : : : : : :
147 GlnPyrIleSerGlnGlyTyrTyAsnThrGluIleThrValLysGlnThr 163
      : : : : : : : : : : : : : : : : : : : : : : : :
542 TATTATCTCAAGCGAGGATATTTTCGCATCCAGGTGTAGACTACAGT... 586
      : : : : : : : : : : : : : : : : : : : : : : : :

```

163 rMetLeuAspGlyAsnArgValLys .....LeuAspMetThrPheAla. 177  
587 ...CTGGAACAACAACTCAAGAAAAAGTCACATCGATGTTTAAATAAAA 632  
178 .....GluGlyLysProAlaArgValValAspIleAsnIleLeGlyAsn 192  
633 TCAATGAAGTCTCTGCGGAAAAATTAACAGGTTTACGTTCTCAGGAATC 682

193 GlnHisPheSerAspAlaAspLeuIleAspValLeuAlaIleLysAspAs 209  
 683 TCTCGATCAGAAAAATCCAGATATCCAGAATTATTTCAAACCAAGCAGCA 732  
 209 nLysIleAsnPro.....LeuSerLysAlaAspA 219  
 733 CTCTTCAACATCAAGTTGGTTTCTGGAGCTGACGTCTATATCACCAGATA 782

219 rgTyrThrGlnGluLysLeuValThrSerLeuCluAsnLeuargAlaLys 235  
 |||:::||||: :::  
 783 TTGTGACACAGATAGCTGGCAATTACG.....AAT 814

[illegible]

```

1667 AACAGAGCATATCTAAGATTATGTCCTCAACCTATGGCGGAACGT 1716
519 uSerTyrGlyTyrProIleAspGluAsnGlnArgIleSerPheGlyLeuA 536
1717 CACACACAGCTATATCTTGAAACCAACCTGAATACGGTCTATTTTAT 1766
536 snAlaAspAsnThrLysLeuHisGlyGlyArg...PheMetGlyIleSer 551
1767 GAGGAAGTCAACAGAGTTTACATGAAACAGTAAGTTCTCTCTAGGGCCA 1816
552 AsnValLysGlnLeuMetAlaAspGlyGlyLysIleGlnValAspAsnAs 568
1817 AATATA.....GACAGCAATAAAGGATTGCTCTCTGCTGC 1851
568 nGlyIleProAspPheLysHisAspTyrThrTyrAsnAlaIleLeuG 585
1852 AGGTGTC.....A 1859
585 lyTrpAsnTyrSerSerLeuAspArgProValPheProThrGlnGlyMet 601
1860 ACTTGAATTAGCATCTCTGATAGTCTCTAGAACTCCAACTACAGGAT 1909
602 SerHisSerValAspLeuThrVal...GlyPheGlyAspLysThrHis.. 616
1910 CGCGGGGGGTGACTTTTGGAGTTCTGCTTTGGAGAACTATCATTT 1959
617 GlnLysValValTyrGlnGlyAsnIleTyrArgProPheIleLysLys 633
1960 TCAAAACACTCTTTAAACAGCTCTATCTATAGAACTTACGCGTAAAG 2009
633 erValLeu.....ArgGlyTyrAlaLysLeu.....GlyTyrGlyAsn 645
2010 GTATTTGAAATCAAAAGGAGCTCAATTATTAACCTATAGCAAT 2059
646 AsnLeu.....ProPheTyrGluAsnPheTyrAlaGlyGlyTy 658
2060 ACTACAGCTGAAGGAGATCTCTGATGAGCGCTTCTCTAGTGGAGA 2109
658 rGlySerValArgGlyTyrAspGlnSerSerLeuGlyProArgSerGlnA 675
2110 GACTACAGTCGGGATATAAATCTTTTATTCGTGTCCTCAAAATCTCG 2159
675 laTyrLeuThrAlaArgArgGlyGlnGlnThrLeuGlyGluValVal 691
2160 CT.....ACAGAACCTCAG 2173
692 GlyGlyAsnAlaLeuAlaThrPheGlySerGluLeuIleLeuProLeuPr 708
2174 GGAGGACTCTCTCGCTCCTTATTTCAGAGAGTTTCAATACCCCTC... 2221
708 opelysGlyAspTrpIleAspGln.....ValArgProValIlePheI 723
2222 .....ATCAGACAACCTAATATTAGTGCCTTTGTATCT 2255
723 leGluGlyGlyGlnValPheAspThrThrGlyMetAspLysGlnThrIle 739
2256 TAGACTCAGGTTTCTC.....GGTTTACAAGAGTATAAGATT 2293
740 AspLeuThrGlnPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLy 756
2294 TCGTTA..... 2299
756 saLaalaAsnArgProLeuLeuThrGlnAspLysGlnLeuArgTyrSera 773
2300 .....AAGACTCTACGTAGTAGTG 2318
773 laGlyValGlyAlaThrTrpTyrThrProIleGly...ProLeuSerIle 788
2319 CTGGATTGGTCTCGGCTCGATGTAATGTTCTCTGTATGTTA 2368
789 SerTyrAlaLysPro.....LeuAsnLysLysGlnAs 799

```

```

2369 GGATTTGGTGGCCCTTCGTCACACGAGACTTTGAATGGAGAAAAAT 2418
799 naspGlnThrAspThrValGlnPheGlnIleGlySerValPhe 813
2419 TGAATGATCTACGAGATTCTTCTTTAGGGGCGCATGTTTC 2461

```

seq\_name: /SID51/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:AAZ01425

seq\_documentation\_block:

ID AAZ01425 standard; DNA; 1038602 BP.

XX AAZ01425;

XX 07-OCT-1999 (first entry)

XX Complete genome sequence of Chlamydia trachomatis.

XX Vaccine; eye disease; conventional trachoma; nonendemic trachoma;  
 KW paratrachoma; inclusion conjunctivitis; genital disease; perihhepatitis;  
 KW nongonococcal urethritis; epidymitis; cervicitis; salpingitis;  
 KW bartholinitis; pneumopathy; venereal lymphogranulomatosis; ss.

OS Chlamydia trachomatis.

XX WO9928475-A2.

XX 10-JUN-1999.

XX 27-NOV-1998; 98WO-IB01939.

XX 04-NOV-1998; 98US-0107077.

XX 28-NOV-1997; 97FR-0015041.

XX 17-DEC-1997; 97FR-0016034.

XX (GEST ) GENSET.

XX Griffais R;

XX WPI; 1999-371125/31.

XX Genome sequence of Chlamydia trachomatis

XX Claim 1; Page 373-656; 1755pp; English.

XX The present sequence represents the complete genome of Chlamydia  
 CC trachomatis. Open reading frames (ORFs) of the genome encode  
 CC polypeptides AY36754-Y37949. The polypeptides can be used as vaccines  
 CC against Chlamydia trachomatis. Antisense and ribozyme sequences can also  
 CC be used to control growth of the microorganism. Chlamydia trachomatis is  
 CC responsible for a large number of diseases, e.g. eye diseases such as  
 CC conventional trachoma, nonendemic trachoma, paratrachoma, and inclusion  
 CC conjunctivitis; genital diseases such as nongonococcal urethritis;  
 CC epidymitis; cervicitis, salpingitis, perihhepatitis, bartholinitis;  
 CC pneumopathy in breast feeding infants; and venereal  
 CC lymphogranulomatosis. The polypeptides of the invention may be of use in  
 CC treating these diseases.

XX Sequence 1038602 BP; 304265 A; 214645 C; 214259 G; 305001 T; 432 other;

alignment\_scores:

Quality: 386.50 Length: 907

Ratio: 0.880 Gaps: 35

Percent Similarity: 48.401 Percent Identity: 21.058

alignment\_block:

US-09-701-711-2 x AAZ01425/rev ..

Align seg 1/1 to reverse of: AAZ01425 from: 1 to: 1038602

14 MetThrMetAlaValMetMetValMetSerThrHisAlaGlnAlaAs 30

206407 CTGCNACTCGCTGTTTACTGTGCTCACC..... 206378

30 pPheMetAlaAsnAspIleThrIleThrGlyLeuGlnArgValThrIleG 47  
||| : : : : : ||| : : : : : ||| : : : : :  
206377 .TTTTCACAAAGTCTTCTGTTCAACTTCAGAGGACGTATGGTCGTAG 206329  
47 luSerLeu.....GlnSerVal 52  
||| : : : : : ||| : : : : : ||| : : : : :  
206328 AGTCTATCACCATTACGACTCAAGGAGAGAACTCAAAATAAACGAGCT 206279  
53 LeuPro.....PheArgLeuGlnValSerGluAsnGlnLe 66  
: : : : : ||| : : : : : ||| : : : : :  
206278 ATTCTTAATAAACAAGCAAGGACGTTGTTCTCAAGCAGATTT 206229  
66 uAlaAspGlyValLysAlaLeuThrGlyAsnPheSerAspValG 83  
: : : : : ||| : : : : : ||| : : : : :  
206228 TGATGAAGATCTAAGAACACTT.....TCGAAAAATTTGATCGAGTAG 206185  
83 ln.....ValTyrHisGlnGluGlyArgIleIleTyrGln.....Val 95  
: : : : : ||| : : : : : ||| : : : : :  
206184 AGCCTATCGTAGAGTTTCGTAATGGACAAGCTGTGATCTCTGATTCTG 206135  
96 ThrGluArgProLeuIleAlaGluIleAsnPheGluGlyAsnArgLeuI 112  
||| : : : : : ||| : : : : : ||| : : : : :  
206134 ACGCAAAATCTGTTATCAGAGATCAATATTTCAGAAATGAAGCTAT 206085  
112 eProLysGluGlyLeuGlnGluGlyLeuLysAsnAlaGlyLeuAlaValG 129  
||| : : : : : ||| : : : : : ||| : : : : :  
206084 CCCACTCATAAATTT..... 206069  
129 lyGlnProLeuLysGlnAlaThrValGlnMetIleGluThrGluLeuThr 145  
||| : : : : : ||| : : : : : ||| : : : : :  
206068 .....CTGAAA.....ACTTTAGAGCTTTATAAAATGATCTTTT 206033  
146 AsnGln.....TyrI 149  
: : : : : ||| : : : : : ||| : : : : :  
206032 GATCGGGAATTTCTTTAAAAATTTTGATCGCGTAAGAACTCTTTATTT 205983  
149 eSerGlnGlyTyrTyrAsnThrGluIleThrVal..... 160  
: : : : : ||| : : : : : ||| : : : : :  
205982 GAAACGAGGTACTACGATTCTCAACTCTCTCTATTCATAATCATAATG 205933  
161 ..LysGlnThrMetLeuAspGlyAsnArgValLysLeuAspMetThrPhe 176  
||| : : : : : ||| : : : : : ||| : : : : :  
205932 AGAAAGAGGCTTTATCGAT.....ATTTCATCGAGATT 205898  
177 AlaGluGlyLysProAlaArgValValAspIleAsnIleIleGlyAsnG 193  
||| : : : : : ||| : : : : : ||| : : : : :  
205897 AAGAAGGAGCTCAGCTCCGATAAAAAATTAACGATTTCCGGGAATAC 205848  
193 nHisPheSerAspAlaAspLeuIleAspValLeuAlaIleLysAspAsnL 210  
: : : : : ||| : : : : : ||| : : : : :  
205847 CGCAACAGAGCATCAGACTTAGTGACATGTTTAACTAAACAATACT 205798  
210 ysIleAsnPro.....LeuSerLysAlaAspArgTyrThrGlnGluLys 224  
: : : : : ||| : : : : : ||| : : : : :  
205797 CCACAACAAGAGGTGTTGTTCACTGGTCCGGAGTGTATCATCCGACATG 205748  
225 LeuValThrSerLeuGluAsnLeuArgAlaIlystYrLeuAsnAlaGlyPh 241  
: : : : : ||| : : : : : ||| : : : : :  
205747 GTAGACAAGACTTTTGTGATATCACAATACTTCCAAAATAAAGGATA 205698  
241 eValArgPheGluIleLysAspAlaLysLeuAsn.....IleAsnGluA 256  
: : : : : ||| : : : : : ||| : : : : :  
205697 TGCT.....GATGCTAAAGTAAGCAAGAGGTCTCTACAG 205663  
256 sp...LysAsnArgIlePheValGluIleSerLeuHisGluGlyGluGln 271  
||| : : : : : ||| : : : : : ||| : : : : :  
205662 ATGCTAAAGGAAACATTACTTTGCTTATCGTTGTAGACAAAGGACCTTTA 205613  
272 TyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAl 288  
||| : : : : : ||| : : : : : ||| : : : : :  
205612 TACACATTAGGTACGTACATATAGAGGATTTCACAGCGTTATCCAAAAG 205563  
288 aGluLeuGluAlaLeuLeuLysPheLysAlaGluGluGlyPheSerGlnA 305  
||| : : : : : ||| : : : : : ||| : : : : :  
205562 ACTCTCGATAAACAACTATTGGTTGGACCTAACTCCTTATATGCCCCAG 205513  
305 laMetLeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAspAsp 321  
: : : : : ||| : : : : : ||| : : : : :  
205512 ATAAATTTGGACTGGAGCACAAAAGATTCTAGCGCTAGCTAGATAT 205463  
322 GlyTyrTyrTyrAlaGlnIleArgProValThrArgIleAsnAspGluSe 338  
||| : : : : : ||| : : : : : ||| : : : : :  
205462 GGCTACGTG.....ValArgArgIle 358  
: : : : : ||| : : : : : ||| : : : : :  
338 rArgThrValAspValGluTyrTyrIleAspProValHisProValTyr 354  
: : : : : ||| : : : : : ||| : : : : :  
205451 CACTAACGTTGATCTCTCTCTCAGCGCACCCCACTCTACCTGTTTACG 205402  
355 .....ValArgArgIle 358  
: : : : : ||| : : : : : ||| : : : : :  
205401 ATGTTACTATFCGAGTGAGTGAGGATCTCCTACAAAATCGGGTTAAT 205352  
359 AsnPheThrGlyAsnPheLysThrGlnAspGluValLeuArgGluMe 375  
: : : : : ||| : : : : : ||| : : : : :  
205351 AAAATCAAGGAACACTCATCTAAGCATGATGTGATTTGCATGAGAC 205302  
375 tArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnLeuSerA 392  
: : : : : ||| : : : : : ||| : : : : :  
205301 TAGCTTTTCCCTCGAGACACTTTTATAGATTAAACACTAGAAGATACAG 205252  
392 rgAlaArgLeuMetArgThrGlyPhePheLysHisValThrValAspThr 408  
: : : : : ||| : : : : : ||| : : : : :  
205251 AGACTCTTTTACGCAACACCGGCTACTTTAAAGTGTAAAGTGTCTATACG 205202  
409 .....ArgProValProAsnSerProAspGlnValAspVa 420  
||| : : : : : ||| : : : : : ||| : : : : :  
205201 GTTCGTTCCCAATTAGATCTCTCGATTCTAACGACCTTTATCGAGATGT 205152  
420 laAsnPheValValGluGluGlnProSerGlySerSerThrIleAlaAlaG 437  
: : : : : ||| : : : : : ||| : : : : :  
205151 TTTTATGAAGTCAAGAGACTGAAACAGGAATCTTGGGCTATTCTTAG 205102  
437 lyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSerGlnAsn 453  
||| : : : : : ||| : : : : : ||| : : : : :  
205101 GATTCACTCCATTGACCATTTATTGAGGAGCAGAAATTCAGAAAGC 205052  
454 AsnPhe..... 455  
||| : : : : : ||| : : : : : ||| : : : : :  
205051 AACTTTGATTATTGGAGCCCGAACTTTCTCAAAAAGGATTCAAATC 205002  
456 ....MetGlyThrGlyLysHisValAsnAlaSerPheSerArgSerGluT 471  
||| : : : : : ||| : : : : : ||| : : : : :  
205001 TTTAAGAGGTGGTGGAGAATACCTCTTCTTAAAGCTAATTTAGGAGATA 204952  
471 hrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheThr..... 485  
||| : : : : : ||| : : : : : ||| : : : : :  
204951 AGGTCAACGATTACACTGTTAAATGGAGCAACCACTCTCTTAAATACC 204902  
486 .....ValAsnGlyValSerLeuSerGlnSerGlyTyrTyrArgLysTh 500  
: : : : : ||| : : : : : ||| : : : : :  
204901 CCTTGGATTCTGGAGTAGAATTAGT..... 204875  
500 rLysTyrAspAsnLysAsnIleSer...AsnTyrValLeuAspSerTyrG 516  
||| : : : : : ||| : : : : : ||| : : : : :  
204874 .AAATCAATTAATAAGCTTTATCAAAAGACTACTCTGTGGATACCTATG 204826  
516 lyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSer 532  
||| : : : : : ||| : : : : : ||| : : : : :  
204825 GAGGGAATATCAGTACCACCTACATTTCTT.....AACGATAAGTTAAA 204782  
533 PheGlyLeu.....AsnAlaAspAsnThrLysLeuHisGlyArgPh 547  
: : : : : ||| : : : : : ||| : : : : :  
204781 TATGGGATGTATTACCGTGGTGCCAAACAAAGCTTAAGTTTGGCGCAAAA 204732  
547 eMetGlyIleSerAsnValLysGlnLeuMetAlaAspGlyGlyLysIleG 564

```
... |||||
204731 AAGTCCAGCTCTAATAGACTTGGACCAGATTAGACAGTAATAAAGGAT 204682
|||:|||||
564 lnValAspAsnAsnGlylleProAspPheLysHisAspThrThrTyr 580
|||||
204681 TCGTTCCGACGGGACTC..... 204662
581 AsnAlaIleLeuGlyTrpAsnTrpSerSerLeuAspArgProValPhePr 597
|||||:|||||:|||||:|||||:|||||
204661 AATGTTCTC.....TATGATTCTATTGATANTCCTAGAAACC 204624
597 othrGlnGlyMetSerHisSerValAspLeuThrValGlyPheGlyAspL 614
|||||:|||||:|||||:|||||:|||||
204623 TACTATGGGAATCCGACGCTTCTTAAACTNTGAATTATCTGTTAGGCG 204574
614 ysThrHisGln.....LysValValTyrGlnGlyAsnIleTyrArgPro 628
|||||:|||||:|||||:|||||:|||||
204573 GAACCTTACCAATTTAGTAACCTAGCAGTAGTGGTTCTATCTATCCTTA 204524
629 PheIleLysLysSerValLeuArgGlyTyrAlaLysLeuGlyTyr..... 643
|||:|||||:|||||:|||||:|||||
204523 TTAACATAAAAGGTGTTTGAAGNCCGTCAGAAAGCTNAGTTTATCAA 204474
644 .....GlyAsnAsnLeuProPheTyrGluAsnPhe 654
204473 ACCTTTCGGAACAACACTGCACAGGCATCTCTGTACGAAAGCTTCT 204424
654 yrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSerSerLeuGly 670
|||||:|||||:|||||:|||||:|||||
204423 TCTAGAGGTGAACCACTGTTCGGGGTTACAAACCTTTATTTATGGA 204374
671 ProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnGlnThrThrLe 687
|||||:|||||
204373 CGGAAGTTTCTCTCT..... 204359
687 uGlyGluValValGlyGlyAsnAlaLeuAlaThrPheGlySerGluLeu 704
|||||
204358 .ACTGAACCAACAGGAGGCTG.....TCTTCCTCTAC 204328
704 leuLeuProLeuProPheLysGlyAspTrpIleAspGln.....ValArg 718
|||||:|||||:|||||
204327 TATTACAGAAGAATTTCAATATCTCTTTCATCTTCACCTTGCAATAT 204278
719 ProValIlePheIleGluGlyGlyGlnValPheAspThrThrGlyMetAs 735
|||||:|||||:|||||:|||||
204277 GCCTTTGTATTCTAGATTCCGATTCAAT.....GGGATAGA 204240
735 pLysGlnThrIleAspLeuThrGlnPheLysAspProGlnAlaThrAlaG 752
|||||:|||||:|||||
204239 AGAGCACATATTCGCTG..... 204221
752 luGlnAsnAlaLysAlaAlaAsnArgProLeuLeuThrGlnAspLysGln 768
|||||:|||||:|||||
204220 .....AAGAC 204215
769 LeuArgTyrSerAlaGlyValGlyAlaThr..... 778
|||||:|||||:|||||
204214 CTTTCAGTAGCGCTGGATTTGGTTGTCTACGCTTTGATGATGAATAATCT 204165
779 .....TriTyrThrProIleGlyProLeuSerIleSerT 790
|||||:|||||:|||||:|||||
204164 GCAATATGCTAGCTGGGTTGGCGGTTCGCCCAACAGAAATC.... 204119
790 yrAlaLysProLeuAsnLysLysGlnAsnAspGlnThrAspThrValGln 806
|||||:|||||:|||||:|||||
204118 .....CTCAATATGAAAAATTTGATGTATCTCAAGAAATCTTTC 204080
807 PheGlnIleGlySerValPhe 813
|||||:|||||:|||||:|||||
204079 TTTGCTTTGGGAGGATATTC 204059
```

seq\_name: /SIDSl/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.AAX91587

```
seq_documentation_block:
ID AAX91587 standard; DNA; 2697 BP.
XX
AC AAX91587;
XX
DT 25-AUG-1999 (first entry)
XX
DE Porphyromonas gingivalis protein PG36 ORF encoding DNA.
XX
KW Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
KW vaccine; antigenic; ds.
XX
OS Porphyromonas gingivalis.
XX
PN WO9929870-Al.
XX
PD 17-JUN-1999.
XX
PF 10-DEC-1998; 98WO-AU01023.
XX
PR 04-AUG-1998; 98AU-0005028.
PR 10-DEC-1997; 97AU-0000839.
PR 31-DEC-1997; 97AU-0001182.
PR 30-JAN-1998; 98AU-0001546.
PR 10-MAR-1998; 98AU-0002264.
PR 09-APR-1998; 98AU-0002911.
PR 23-APR-1998; 98AU-0003128.
PR 05-MAY-1998; 98AU-0003338.
PR 22-MAY-1998; 98AU-0003654.
PR 29-JUL-1998; 98AU-0004917.
XX
PA (CSLC-) CSL LTD.
XX
PI Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
PI Ross BC, Rothen LJ, Webb EA;
XX
DR WPI; 1999-385613/32.
DR P-PSDB; AAY34369.
XX
XX Antigenic Porphyromonas gingivalis peptides for preventing
gingivitis
XX
PS Claim 12; Page 129-130; 588pp; English.
XX
XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic
porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
isolation of the PG polypeptides. The PG polypeptides have antibacterial
activity with a vaccine mechanism of action. The PG polypeptides can be
used as vaccines especially against Porphyromonas gingivalis. Probes can
be used to detect Porphyromonas gingivalis in standard hybridisation
assays. Porphyromonas gingivalis is involved in periodontal disease
especially gingivitis.
XX
SQ Sequence 2697 BP; 757 A; 551 C; 642 G; 747 T; 0 other;
```

```
alignment_scores:
Quality: 379.50 Length: 976
Ratio: 0.794 Gaps: 43
Percent Similarity: 48.975 Percent Identity: 20.799
alignment_block:
US-09-701-711-2 x AAX91587 ..
Align seg 1/1 to: AAX91587 from: 1 to: 2697
2 ArgAsnSerTyrPheLysGlyPheGlnValSerAlaMetThrMetAlaVa 18
: : : : : : : : : : : : : : : : : : : : : : : : : : :
19 AAGACTATGTACAAAAAGATTATTGCGGTAGCAGCTCTCTCTGCGCAG 68
: : : : : : : : : : : : : : : : : : : : : : : : : : :
18 lMetMetValMetSerThrHisAlaGlnAlaAlaAspPheMetAlaAsnA 35
: : : : : : : : : : : : : : : : : : : : : : : : : : :
```

[illegible]

545 lYArgPheMetGlyIleSerAsnValLysGlnLeuMetalAlaAspGlyGly 561  
1837 ..CAGATGTTAGTACTTCG.....ATCGTTACGGTAAG 1869  
562 LysIleGlnValAspAsnGlyIleProAspPheLysHisAspTyrTh 578  
1870 CGTTGACTTGGCGGACAAATGGTTCCAGATTATATCTTCTCGAAGTA 1919  
578 rThrTyrAsnAlaIleLeuGlyTyrAsnTyr..... 588  
1920 CACCTACTATAGATGCGAATATGGAGCTACAATACCTTCCAAATTTCC 1969  
589 .....Ser 589  
1970 ATCATGGCTCGGCTAAATGATCTCAACTTGGAGCTGCGTCTCTCGTACT 2019  
590 SerLeuAspArgProValPheProThrGlnGlyMetSerHisSerValAs 606  
2020 TCCATCGATAATCCTATTATACAGAAAGCGGATTCATGGTTTC 2069  
606 pLeu..... 607  
2070 TGTGCTGCTACTCTCTTATTCTTTGGGACAATCATGACTATGCCA 2119  
608 .....ThrValGlyPheGlyAspLys.....ThrHisGlnLys 618  
2120 GCCAGAACCTCAGCGTAAGCGATGTTACAGATTATCGATATCACAAAG 2169  
619 ValValTyrGlnGlyAsnIleTyrArgProPheIleLys..... 631  
2170 TGGAGTTTAGAGGACGAGTTTACTCCATTGCTCAATCTGCTACGCCA 2219  
632 .....LysSerValLeuArgGlyTyrAlaLysLeuG 642  
2220 TAAATATACACCGTGCATCAGTGCAGTGGAGGACAGTCTTGTGTT 2269  
642 lyTyrGlyAsnAsn.....LeuProPheTyrGluAsnPheTyrAlaGly 656  
2270 CGTAAATTCCAATAAGAAATCTCCTTTC...GGTACTTTCTATATGGGA 2316  
657 Gly.....TyrGlySerValArgGlyTyrAspGlnse 667  
2317 GGTGATGATGTCCAGCTATTATGGT.....GGCTACATCAATGA 2357  
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684  
2358 GACTATAGGT.....TTGCGTGGTTATA 2380  
684 lnThrThrLeuGlyGluValValGlyGlyAsn.....Ala 695  
2381 AGAAC.....GGATCTATTCGGGTAATAACTACGACTATGCATATGCT 2424  
696 LeuAlaThrPheGlySerGluLeuLeuProLeuProPheLysGlyAs 712  
2425 TATATGCGCTTACGATGGAACACTACGTTCCCGATTCTGTTGAAACTC 2474  
712 p.....TrpIleAspGlnValArgProValIlePheIleGluGly 726  
2475 ATTCAATGCGTGGCTC.....TTAGCTTTTCCCGAAGCAG 2509  
726 GlnGlnValPheAspThrThrGlyMetAspLysGlnThrIleAspLeuThr 742  
2510 GCAATGCGGTGG.....CGCAGTATCGAC..... 2532  
743 GlnPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAs 759  
2533 .....AATTATATCCCTTTAA 2549  
759 nArgProLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValG 776  
2550 C.....CTGAAGCATCGCGCGGTGTAG 2572  
776 lyAlaThrTrpTyrThrPro...IleGlyProLeuSerIleSerTyrAla 791

2573 GATTGCGTGTAAACGTTACGATGGTCCGAATGCTCGGTATCGATTGGGA 2622  
792 .....LysProLeuAsnLysLysGlnAsnAspGlnThrAspTh 804  
2623 TATGCGTTTGACCGTCCGCAAAATCTCTACAGCGAGGAGGAAGCAAT... 2670  
804 rValGlnPheGlnIleGlySerValPhe 813  
2671 .GTCCACTTTGCTCGCAGCAGAGTTTC 2697  
seq\_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.AAX91713  
seq\_documentation\_block:  
ID AAX91713 standard; DNA; 2673 BP.  
XX  
AC AAX91713;  
XX  
DT 25-AUG-1999 (first entry)  
XX  
DE Porphyromonas gingivalis protein PG36 encoding DNA.  
XX  
KW Porphyromonas gingivalis; PG; periodontal disease; gingivitis;  
XX vaccine; antigenic; ds.  
XX  
OS Porphyromonas gingivalis.  
XX  
PN WO9929870-A1.  
XX  
PD 17-JUN-1999.  
XX  
PF 10-DEC-1998; 98WO-AU01023.  
XX  
PR 04-AUG-1998; 98AU-0005028.  
XX 10-DEC-1997; 97AU-0000839.  
PR 31-DEC-1997; 97AU-0001182.  
PR 30-JAN-1998; 98AU-0001546.  
PR 10-MAR-1998; 98AU-0002264.  
PR 09-APR-1998; 98AU-0002911.  
PR 23-APR-1998; 98AU-0003128.  
PR 05-MAY-1998; 98AU-0003338.  
PR 22-MAY-1998; 98AU-0003654.  
PR 29-JUL-1998; 98AU-0004917.  
XX  
PA (CSLC-) CSL LTD.  
XX  
PI Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;  
PI Ross BC, Rothel LJ, Webb EA;  
XX  
DR WPI; 1999-385613/32.  
DR P-PSDB; AAY34495.  
XX  
PT Antigenic Porphyromonas gingivalis peptides for preventing  
PT gingivitis  
XX  
PS Claim 12; Page 215-216; 588pp; English.  
XX  
CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic  
CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to  
CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the  
CC isolation of the PG polypeptides. The PG polypeptides have antibacterial  
CC activity with a vaccine mechanism of action. The PG polypeptides can be  
CC used as vaccines especially against Porphyromonas gingivalis. Probes can  
CC be used to detect Porphyromonas gingivalis in standard hybridisation  
CC assays. Porphyromonas gingivalis is involved in periodontal disease  
CC especially gingivitis.  
XX  
SQ Sequence 2673 BP; 747 A; 549 C; 638 G; 739 T; 0 other;

alignment\_scores:  
Quality: 378.50 Length: 973  
Ratio: 0.795 Gaps: 43



Percent Similarity: 48.921 Percent Identity: 20.863

alignment\_block:

US-09-701-711-2 x AAX91713

Align seg 1/1 to: AAX91713 from: 1 to: 2673

```
5 TyrPheLysGlyPheGlnValSerAlaMetThrMetAlaValMetMetVa 21
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4 TACAAAAGATTATTCGCGTAGCAGCTCTCTCTCGCGCCAGCATAGGGAT 53
21 lMetSerThrHisAlaGlnAlaAlaAspPheMetAlaAsnAlaLeu... 36
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
54 CCTGAAA.....GGACAGCTCTCGGATCTGACCCCTCAGGATACATAT 97
37 .....ThrIleThrGly 40
98 ATAGCCCTGAATATCCTATGCAAGCCCTATTCATAAGACCATGATCT 147
41 LeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPheArgLe 57
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
148 ATTGAG.....ATCGAGGATGAGGCTTTTCGATGACTTTGTCTT 188
57 u.....GlyGlnValValSerGluAsnGlnLeuAlaAspGly. 69
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
189 GCGCAATCTTTTCAGGCTTGGCTGTAGGTGATGAAGTCTGATTCCTGGAG 238
70 .....VallysAlaLeuTyrAlaThrGlyAsnPheSer 80
239 ATGCCATCTGCTGCTGCGTGAATGAATATTCGCGAGGCTACTCTCA 288
81 AspValGlnValTyrHisGlnGlu.....GlyArgIleIleTyrGln... 94
289 AATGTGCAATCATCGCGGATAAATATCTCGCATAAAGTCTATCTGAA 338
95 .....ValThrGluArgProLeuIleAlaGluIleAsnPheGluGlyA 109
339 AATCATTTCTCACTGAACCTCTCGCATCAGTAAGGTTACTTTTAGCGGG 388
109 snArgLeuIleProLysGluGlyLeuGlnGluGlyLeuLysAlaGly 125
389 TAAGAAGCTCGAGAGAGATCTTGAA.....ATGAAATCGGT 429
126 LeuAlaValGlyGlnProLeuLysGlnAlaThrValGlnMetIleGluTh 142
430 CTTCCGAGGGGATTTCAGATCACCAGAAATATGAAGACAAAGTTCAGGCA 479
142 rGluLeuThrAsnGlnTyrIleSerGlnGlyTyrTyrAsnThrGluIleT 159
480 AATCGTACAGAAGTATTTTAGTGAGAAAGTTATCGCGATGCGCATAC 529
159 hrValLysGln.....ThrMetLeuAspGlyAsnArgValLysLeuAsp 173
530 GGATAACCGACGACCGGATCTTTCCAAAGATGGCTTTGTCATGTGCTT 579
174 MetThrPheAlaGluGlyLysProAlaArgValValAspIleIleI 190
580 ATCTCGATTGAGAGAAAGCAAAACCAAGGTGAATGAATTTATTTTC 629
190 eGlyAsnGlnHisPheSerAspAlaAspLeuAspValLeuAlaIleL 207
630 CGGCAACCAAGCCCTTAGCAATCATAAAGCTA.....AGAATGGCGGATGA 673
207 ysAspAsnLysIleAsn..... 212
674 AGAACCCCAATGCCAAATTCAGTCTTAGAAGCATATTCGCTCATCTTTC 723
213 ...ProLeuSerLysAlaAspArgTyrThrGlnGluLysLeuValThrSe 228
724 TTGAAACTTTTATGACTACTCATAGTTTCTGGAGAGAGACCTACCGTGAAGA 773
228 rLeuGluAsnLeuArgAlaLysTyrLeuAsnAlaGlyPheValArgPheG 245
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
774 TTTGGTCCGATTGATAGAGAAGTATCAGGAATATGATATCGTATCGTCTG 823
245 luIleLysAspAlaLysLeuAsnIleAsnGluAspLysAsnArgIlePhe 261
824 AATACTGACCGACAGTGTCTGTAAGGCTCTGACGCGCAAGAGTGGAT 873
262 ValGluIleSerLeuHisGluGlyGlnTyrArgPheGlyGlnThrGl 278
874 ATTATCTCAACATCGAAGAGGGCAGAAATATATTAAGGATGTCAA 923
278 nPheLeuGlyAsnLeuThrTyrThrGlnAlaGluLeuGluAlaLeuLeu 295
924 CTTTGTGGCAATTCAATATCCATCGGAGTATTTGGAACGAGTCTGTCG 973
295 ysPheLysAlaGluGlyPheSerGlnAlaMetLeuGluGlnThrThr 311
974 GAATAAATCCGGAGATGTACAATCAGAGAGGATTGGCTAAGCGTCTC 1023
312 AsnAsn.....IleSerThrLysPheGlyAspGlyTyrTy 324
1024 AATGAAGATGAAGATCTCTGGGGAACCTGTACTATAACAATGGCTATAT 1073
324 rTyrAlaGlnIleArgProVal...ThrArgIleAsnAspGluSerArgT 340
1074 TTTTGGTGGTCCGATCCGCTGGAACAAATGTAGTGGGGATCTGTTT 1123
340 hrValAspValGluTyrTyrIleAspProValHisProValTyrValArg 356
1124 CCTTGTGATATTCGT.....ATAGCGGAGGGGAAGCAGGCCAATATCAAT 1167
357 ArgIleAsnPheThrGlyAsnPheLysThrGlnAspGluValLeuArg 373
1168 AAGGTGATCATCAAGGAATAACTGTCGTGACGAAGACGTAGTACGCG 1217
373 gGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnL 390
1218 AGAGCTTTACACAAAGCCGCGCAGCTCTTAGTCGCGAGGATATCATTA 1267
390 euSerArgAlaArgLeuMetArgThrGlyPhePheLysHisValThrVal 406
1268 ACTCTATTCTCATCAATCAGCTTGGGCATTTTCATGCCGAAAATCT 1317
407 AspThrArgProValProAsn...SerProAspGlnValAspValAsnPh 422
1318 ATTCCTCCGTCGATTCCTCAATCCCGAAACAGGAACAGTGGATAGAGTA 1367
422 eValValGluGlnProSerGlySerThrIleAlaAlaGlyTyrS 439
1368 TGATTTGGTCCGCGTAGCAGTGAACCAATGGAGCTTTCTGTCGGTTGGA 1417
439 erGlnSerGlyGlyValThrPheGlnPheAspValSerGlnAsnAsnPh 455
1418 GTCACTCC...GGACTTCTGTCGAGGAGGCCATTAAGTTCACGAACCTTC 1464
455 ..... 455
1465 TCTGTGCGCAACTTGTCTCCATCCCTCGATGTATAAGAAAGGGATCATTC 1514
456 .MetGlyThrGlyLysHisValAsnAlaSerPhe...SerArgSerGluT 471
1515 GCAAGGGGATGGCAACACATATCACTAGTGTCTCAGACCAATGGAAAGT 1564
471 hrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheThrValAsn 487
1565 ACTATCAGCAGATATGTGTCACATTTATGCGATCCATGGTTTGGGGCAAG 1614
488 GlyValSerGln...SerLeuSerGlyTyrTyrArgLysThr..... 500
1615 CGCGCGGATATGTCAGCTTCAGTGTCTATTCACAGACTACGCGGAT 1664
501 .....LysTyrAspAsnLysAsnIleSerAsnTyrVal..... 511
1665 TGACTCCAAGTTCTACAATAGCAATGCCGCAACTACTATAATGCCTACT 1714
```







SQ Sequence 2439 BP; 596 A; 544 C; 731 G; 568 T; 0 other;

alignment scores:

Quality:	348.50	Length:	847
Ratio:	0.783	Gaps:	28
Percent Similarity:	52.538	Percent Identity:	21.133

alignment block:

US-09-701-711-2 X AAZ33106

Align seq 1/1 to: AAZ33106 from: 1 to: 2439

36 lIeThrIleThrGlyLeuGlnArgValThrIleGluSerLeuGlnSerVa 52  
31 ATTAGTTTGGGGGCTCGAATATATCTCTCGCGCCGATTTGGACACGAT 80  
52 lLeuProPheArgLeuGlyGlnValValSerGluAsnGlnLeuAlaAspG 69  
81 TTTTCTCAATACAGGACAAAGATGACCTATGAGCTGTACCTGGAGA 130  
69 lyVallysAlaLeuTyrlaThrGlyAsnPheSerAspVal. .... 82  
131 TACTGCAAAAGGTCTATGACCTTGATGACTTTCTTGAAGTTTCGCTTAAG 180  
83 .....GlnValTyHisGlnGluGlyArgIleIleTyrlGlnVa 95  
181 GCGGTGCCACCGATCCGGAGTATCAGTATGTGATGTACAGTTCACGGT 230  
95 lThrGluArgProLeuIleAlaGluIleAsnPheGluGlyAsnArgLeuI 112  
231 AAGGAGCGCTCTCGGTGAAGGGCATCAAGATGTTGAAGCAACAGCCAA 280  
112 leProLysGluGlyLeuGlnGluGlyLeuLysAsnAlaGlyLeuAlaVal 128  
281 TCCGCGAGTGGGACCTT.....TTGTCTAAATCTCTCTGAAAG 321  
129 GlyGlnProLeuLysGlnAlaThrValGlnMetIleGluThrGluLeuTh 145  
322 GGAGACATTTACAATGAAGTAAAGATGAAGTGGACCAAGAGTCGCTCAG 371  
145 rAsnGlnTyrlIleSerGlnGlyTyrlTyrlAsnThrGluIleThrValLysG 162  
372 GCGTCATTACCTGGACAGGGCTATGCGCGGGTTTAAGATATCTCTGC...G 418  
162 lInThrMetLeuAspGlyAsnArgValLysLeuAspMetThrPheAlaGlu 178  
419 AGCAAAACTGAGCGGGGGCGTGGTGTCACAGTTTACCATTCAGGAA 468  
179 GlyLysProAlaArgValValAspIleAsnIleIleGlyAsnGlnHisPh 195  
469 GGTAAACAGACATGTCTCTCGCGGATACAGTTTAAAGGAAATAAGCGCT 518  
195 eSerAspAlaAspLeuIleAspValLeuAlaIleLysAspAsnLysIleA 212  
519 TACCGAGTCGGTCTCAGAGAGGTGCTTTCACCGCAGGAGCGCGT... 564  
212 snProLeuSerLysAlaAspArgTyrlThrGlnGluLysLeuValThrSer 228  
565 ....TTTTTGACCATGGGTGTTCAAGGAGAAATGCGCTGGAAGCGGAT 609  
229 LeuGluAsnLeuArgAlaLysTyrlLeuAsnAlaGlyPheValArgPheG1 245  
610 AAGCGGCGAGTCCACTCATATCTAGACAGAGGGGATACATTGACGCGCG 659  
245 uIleLysAspAlaLysLeuAsnIleAsnGlu.....AspLysA 258  
660 GTTAGAGCGGTGGCAAGACGGTTGATAAAAAACTGACCCAGTCGCGA 709  
258 snArgIlePheValGluLysSerLeuHisGluGlyGluTnTyrlArgPhe 274  
710 ATCTGGTACCGCTTACGTACACTGTGGTGAAGGTAGACAGTACCGCTAC 759

275 GlyGlnThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAlaGluLeuG1 291  
||| : : : : : ||||| : : : : : ||||| : : : : :  
760 GCGGGGTACCATTGGTGAACAGATTTTTATAGCACCGGAGGCTGCA 809  
291 AlaLeuLeuGlyPheLysAlaGluGluGlyPheSerGlnAlaMetLeuG 308  
: ||| : : : : : ||| : : : : : : : : : : : : : : :  
810 GGCAAAATTAGGTCAAGCGCGGGCCATCATGAATATGTCGCCCTTG 859  
308 LuGlnThrThrAsnAsnIleSerThrLysPheGlyAspSpGlyTyr 324  
||| : : : : : : : : : : : : : : : : : : : : :  
860 AGCAGGCTTTCAGCGCTGCGGATCGGTATTTGAAACGAGTACACG 909  
325 TyrAlaGlnIleArgProValThrArgIleAsnAspGluSerArgThrVa 341  
: : : : : : : : : : : : : : : : : : : : : : : : : :  
910 TCAAAATACCTGAACAAGAAAGAACACGGGACACGCGCAGAAAACGT 959  
341 LaspValGluTyrTyrIleAspProValHisProValTyrValArgArgI 358  
: : : : : : : : : : : : : : : : : : : : : : : : : :  
960 TTCGTTTAAGATACAGGTGGTGAGCGGACGCGGACCGCATCGAGACA 1009  
358 LeAsnPheThrGlyAsnPhelyThrGlnAspGluValLeuArgArgGlu 374  
||| : : : : : : : : : : : : : : : : : : : : :  
1010 TTATCATTAAGGAAACGAAGAATAACAAGACGAGGTATCTCCGCTGAA 1059  
375 MetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnLeuSe 391  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
1060 ATGCTGCTGAACCGGGGATGTCTTCTCAAGTCAAAAGTTTACGGATAG 1109  
391 rArgAlaArgLeuMetArgThrGlyPhePheLysHisValThrValAspT 408  
: : : : : : : : : : : : : : : : : : : : : : : : : :  
1110 CTGCGCAATCTGTCAACCTGCCGTATTCTCGCTCGTGGCGGGATG 1159  
408 hrArgProValProAsnSerProAspGlnValAspValAsnPheValVal 424  
: : : : : : : : : : : ||| ||||| : : : : : : : : : :  
1160 TCGGCCCC...GGCTCTCAGCAGGACCTGGTGGACATATCTCTCAATGTG 1206  
425 GluGluGlnProSerGlySerSerThrIleAlaAlaGlyTyrSerGlnSe 441  
||| ||||| : : : : : : : : : : : : : : : : : : : : :  
1207 GAGGACAGTCCAGCGGAAACGTCAGTTTGGGTGACGTTTCTGGGT 1256  
441 rGlyGlyVal...ThrPheGlnPheAspVal.....SerClnA 453  
||| : : : : : ||||| : : : : : : : : : : : : : : :  
1257 GGGGAGGACGACGTCGCTTTCGCTCTTTGTACGTGGGAAGAAA 1306  
453 snAsnPheMetClyThrGlyLysHisValAsnAlaSerPheSerArgSer 469  
: : : : : : : : : : ||| : : : : : : : : : : : : : : :  
1307 AGAATTTTTCGGAAAGGGAATCAAAATTCAGTAATGCAACCTTGGGG 1356  
470 GluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheThrVa 486  
: : : : : : : : : : ||| : : : : : : : : : : : : : : :  
1357 TCTGAGGCGCAGACCTGGAAGCTCGGTATGTGGAGCGCTGTTTCTGGG 1406  
486 LAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLysTyrA 503  
: : : : : : : : : : : : : : : : : : : : : : : : : :  
1407 CTCCTCCGCTGACGTGGCTTTGAC.....TTTGAACCTTA 1441  
503 spAsnLysAsnIleSerAsnTyrValLeuAspSerTyrGlyGlySerLeu 519  
: : : : : : : : : : ||| : : : : : : : : : : : : : : :  
1442 CGCAAAAAATCTCTTGTGTACCGCGGGTTCATACGCGACACGGGCTG 1491  
520 SerTyrGlyTyrProIleAspGluAsnGlnArgIleSerPheGlyLeuAs 536  
: : : : : ||| : : : : : ||| : : : : : : : : : : : : : :  
1492 CCGCACCGGTACACGACGAGGACGTCGGCTAGTTCCCTCGGGCTG... 1539  
536 nAlaAspAsnThrLysLeuHisGlyGlyArgPhe.....MetGlyI 550  
||| : : : : : : : : : : : : : : : : : : : : : : : : : :  
1540 .GCAGATCGTTTCGCTCAAGTATTCGCGCTTTGAGTCGCGCATCGGCG 1588  
550 LeSerAsnValLysGlnLeuMetAlaaspGlyGlyLysIleGlnValasp 566  
: : : : : ||| : : : : : : : : : : : : : : : : : : : :  
1589 CGCACCGGGTACCACTGGATCCCGGCTATCGGTCATTAGGGTG... 1635  
567 AsnAsnGlyIleProAspPheLys..... 574

```
1636 ...ACCGGGGGGTGACGTTTCGGTTGTAAGAAATTTTACGATAAGGA 1682
575 .....HisAspTyrThrT 579
1683 TAACATCAGCCCTTCAGCTGACCTAACAGAGCAGCTGAACCTGGACCA 1732
579 hrTyrAsnAlaLeuLeuGlyTyrAsnTyrSerSerLeuAspArgProVal 595
1733 GTATCAATTCGTTT.....TGACGAGCGCTTCGTTTCAC..... 1767
596 PheProThrGlnGlyMetSerHisSerValAspLeuThrValGlyPheL 612
1768 .....GGCGGTGACTTCGTTCAGCACCCGCTCCAGCGGCTGG... 1803
612 yAspLysThrHisGlnLysValValGlyGlnGlyAsnIleTyrArgProP 629
1804 ....TTTTAGACAGCGCTGTACGTTCAACGGGCTCGTT.....CCCT 1843
629 heLleLysLysSer...ValLeuArgGlyTyrAlaLys..... 640
1844 TTCTCGAAAAGAGCATTCGTTTCGCTCGCACCAAGCCGCGAGTCTAC 1893
641 .....LeuGlyTyrGlyAsn.....AsnLeuProPheTyr 650
1894 GTTACCTGCTCAATATATCCGCTCTCTCGCGGTGGAACTTAAAGTTTGT 1943
650 rGluAsnPheTyrAlaGlyTyrGlySerValArgGlyTyrAspGlnS 667
1944 CTTGGCTTCTACACCGGT.....GTGCTCGTTCAACCGTAT..... 1980
667 erSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgGlyGln 683
1981 .....TATGGACGGGAAAAAGC 1998
684 GlnThrThrLeuGlyGluValValGlyGlyAsnAlaLeuAlaThrPheL 700
1999 GAAACGGAAGGCAACGGGTGGCGGTCCGGCGCTGGTAATAGACGG 2048
700 y.....Serg 702
2049 CQTGCTGTAGGCGCGGTGGAGGAGAGACGCAAGAAACACCGGAG 2098
702 luLeuIleLeuProLeuProPheLysGlyAspTyrPheAspGlnValArg 718
2099 ACCTGTCTCTC.....CACCACTGGATTGAGTTCGCTGG 2133
719 ProVal.....IlePheLleGlyGlyGly 727
2134 CCGCTGGCGCACCGCATGTGCTCTTTCGACTTTTCTTTGATGGGCAAT 2183
727 nValPheAspThrThrGlyMetAspLysGlnThrIleAspLeuThrGlnP 744
2184 GGTGTACAAACATCGAA.....AGTCAGT 2206
744 heLysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArg 760
2207 CCCCAAGCGGTCTATCGTCCCGCAGCAGCTCCAGCAGCAGCTAGTAGT 2256
761 ProLeuLeuThrGlnAspLysGlnLeuArgTyr.....SerAlaGlyVa 775
2257 AGCAGTAGAACACCAACGACGCTCTGAAGGACTGTACAAAATAGACTACGGTCC 2306
775 lGlyAlaThrTyrThrProIleGlyProLeuSerIleSerTyrAlaL 792
2307 GGGCTGGCTTTACATTGCCGCAATTCGTTTAAATTTGGCTTCGCAA 2356
792 ys.....ProLeuAsnLysLys 797
2357 ACACCTTCACGTCACCGGGCGGCATCCCAAAACAAAGAA 2397
```

seq\_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:AAV25052

seq\_documentation\_block:  
ID AAV25052 standard; DNA; 1845 BP.

XX AC AAV25052;

XX DT 24-JUN-1998 (first entry)

XX DE H. pylori cell envelope OMP ORF 05ep10815\_4719175\_c1\_83.

XX KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
KW identification; binding compound; bacteria; life cycle; activator;  
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis; ds.  
XX OS Helicobacter pylori.XX FH Key Location/Qualifiers  
FT CDS 1..1845  
FT /\*tag= a

XX PN WO9737044-A1.

XX PD 09-OCT-1997.

XX PF 27-MAR-1997; 97WO-US05223.

XX PR 06-DEC-1996; 96US-0761318.

XX PR 29-MAR-1996; 96US-0625811.

XX PR 02-APR-1996; 96US-0758731.

XX PR 25-OCT-1996; 96US-0736905.

XX PR 28-OCT-1996; 96US-0738859.

XX PA (ASTR ) ASTRA AB.

XX PI Alm RA, Smith D;

XX DR WPI; 1997-503122/46.

XX DR P-PSDB; AAW55643.

XX PS Claims 5,6,21; Pages 417-418; 1145pp; English.

This DNA is stated to encode a H. pylori cell envelope outer membrane protein having a terminal Phe residue. The protein may be used in a vaccine to prevent or treat H. pylori infection or to identify H. pylori polypeptide binding compounds, useful as potential H. pylori life cycle activators or inhibitors. The DNA and probes derived from it may be used for the identification of H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic acid sequences complementary to the DNA act as antisense sequences and can be used to prevent the translation of H. pylori mRNA. Antibodies against the protein can be used in immunoassays to evaluate the abundance and distribution of H. pylori-specific antigens. The genomic sequence of H. pylori (ATCC 55679) was determined from overlapping contigs generated by mechanically shearing the bacterial DNA. The sequences were analysed for ORF of at least 180 nucleotides, and the predicted coding regions defined by computer evaluation. To identify likely H. pylori antigens for vaccine development, the amino acid sequences predicted from various ORF were analysed for significant homology to other known or exported membrane proteins. Having identified and determined the sequences of interest, particular regions can be isolated from H. pylori by PCR amplification for recombinant polypeptide production, e.g. in E. coli hosts.

Sequence 1845 BP; 560 A; 329 C; 473 G; 483 T; 0 other;

alignment\_scores:

Quality:	344.50	Length:	557
Ratio:	1.101	Gaps:	17
Percent Similarity:	56.194	Percent Identity:	23.339

alignment\_block:

US-09-701-711-2 x AAV25052

Align seg 1/1 to: AAV25052 from: 1 to: 1845

```

30 AspPheMetAlaAsnAspIleThrIleThrGlyLeuGlnArgValThrIle 46
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
85 GAAATGAAAGTCAAGTCCATTTCTTATGTCGGGCTTTCTTACATGCTGA 134
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
46 eGluSerLeuGlnSerValLeuProPheArgLeuGlyGlnValValSerG 63
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
135 CATGTCGCTATGAATGCAAGATTCGCGGGCGATATGGTGGATT 184
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
63 luAsnGlnLeuAlaAspGlyValLysAlaLeuTyrAlaThrGlyAsnPhe 79
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
185 CTAAAGAAATAGACACCGCTGTTTGTGCTTGTCAACCAAGGGTATT 234
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
80 SerAspValGlnValTyrHsGlnGluGlyArgIleIleTyrGlnValTh 96
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
235 AAAGACGTTTATGCCACTTTTGAACACGGCATTTTAGAGTTTCATTTGA 284
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
96 rGluArgProLeuIleAlaGluIleAsnPheGluGlyAsnArgLeuIleP 113
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
285 TGAAGAGCCGAGATTGCCGGGTGAATATCAAGGT.....TATG 325
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
113 roLysGluGlyLeuGlnGluGlyLeuLysAsn...AlaGlyLeuAlaVal 128
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
326 GGACTGAAAGAGAAAGACCGGCTTAAATCCCAATGGGGATCAAAAG 375
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
129 GlyGlnProLeuLysGlnAlaThrValGlnMetIleGluThrGluLeuTh 145
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
376 GCGACACCTTTGATGAGCAAAATATAGACATGCTTAAACCGGCTTAAA 425
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
145 rAsnGlnTyrIleSerGlnGlyTyrAsnThrGluIleThrValLysG 162
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
426 AACGGCTTTAGAGGCGCGGCTATTATGGAGCGTGGTGGAGGTCGCA 475
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
162 lnThrMetLeuAsp..... 166
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
476 CAGAAAGCTCAGTGAGGAGCGCTTATTGATCCTGTTGATGTAATAGG 525
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
167 GlyAsnArgValLysLeuAspMetThrPheAlaGluGly.....LysPr 181
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
526 GGGACAGTATTATATCAACCAATCCATTATGAGGACGAGCATAAATT 575
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
181 oAlaArgValValAspIleAsnIleIleGlyAsnGlnHisPheSerAspA 198
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
576 AAACCGCGTGTGATTGAATCTTTGAGCGCGAACAAG.....CAGC 616
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
198 laAspLeuIleAsp...ValLeuAlaIleLysAspAsnLysIleAsnPro 213
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
617 GCGATTTTCATGGCTGGATGTCGGGCTTGAATGACCGGAAATG..... 660
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
214 LeuSerLysAlaAspArgTyrThrGlnGluLysLeuValThrSerLeuGl 230
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
661 .....CGCTTAGATCAATTAGAAATACGATTCTTTGCGTATCCAGATGT 704
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
230 uAsnLeuArgAlaLysTyrLeuAsnAla.....GlyPheValA 243
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
705 GTATATGCGTAGGGGTACTTACGCTCATATTTCTTCGCCCTTTTGA 754
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
243 rg.....PheGluIleLysAlaLysLeuAsnIleAsnGluAspLys 257
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
755 AAACGGATTTTCCACCCATGACGCTAAGTCCAT..... 789
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
258 AsnArgIlePheValGluIleSerLeuHisGluGlyGluGlnTyrArgPh 274
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
790 .....TATAAGGTCAAGAGAGGGGATCCAAATACAGGAT 821
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
274 eGlyGlnThrGlnPhe...LeuGlyAsnLeuThrTyrThrGlnAlaGluL 290
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
822 TTCAGATATTTTAAATAGAGATTGACAACCCGGTAGTCCCTTAAAAACCT 871
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

```

```

290 euGluAlaLeuLeuLysPheLysAlaGluGluGlyPheSerGlnAlaMet 306
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
872 TAGAAAAAGCGCTTAAAGTTAAAGGAAGAGATGCTTTTAATTATGAGCAT 921
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
307 LeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAspAspGlyTy 323
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
922 TTAAGAGCGGATCGCAAAATTTTAAAAACCGAAATCGCGATAGGGCTA 971
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
323 rTyrTyrAlaGlnIleArgProValThrArgIleAsnAspGluSerArgT 340
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
972 TCGCTTTGCGTGGTGAAGCCAGACTTGGATAAAGACGAAAAACCGGC 1021
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
340 hrValAspValGluTyrTyrIleAspProValHisProValTyrValArg 356
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1022 TTCTGAAAGTCAATTATCGTATTGAAGTGGCGATATGTTGCATATCAAT 1071
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
357 ArgIleAsnPheThrGlyAsnPheLysThrGlnAspGluValLeuArgAr 373
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1072 GATGTCATCATTTTCAGGGAACCGACGCGATAGGATCATATTAGGAG 1121
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
373 gGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnL 390
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1122 GGAATG.....TTACTAGGCGCTAAAGATAAATACAAC 1156
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
390 euSerArgAlaArg.....LeuMetArgThrGlyPhePhe 401
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1157 TGACCAAACTGAGAAATTCGAAAAATTCCTTTGAGCGCTTTAGCGTTTTC 1206
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
402 LysHisValThrValAspThrArgProValProAsnSerProAspGlnVa 418
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1207 TCTAAAGTCAAGATTGAAGAAAAAAGGGTCAATAGTCA.....TTGAT 1250
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
418 laSpValAsnPheValValGluGlnProSerGlySerSerThrIleA 435
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1251 GGATTTGTTAGTAGCGTAGAAGGGGCGCCAGCGGCGCTGCAATTG 1300
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
435 laAlaGlyTyrSerGlnSerGlyValThrPheGlnPheAspValSer 451
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1301 GGTGGCTATGCTCTTATGAGGGCTCATGCTTAATGGAGCGTGAGC 1350
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
452 GlnAsnAsnPheMetGlyThrGlyLysHisValAsn.....AlaSerPh 466
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1351 GAAAGGAATCTTTTGGCAGGCAAGGCAAGCATGAGCTTGTATGCTAACAT 1400
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
466 eSerArgSerGluThrArgGlu..... 473
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1401 TGCCACAGGGGGGTAGATCTTATCGGGCATGCCAAAAGGGCGGGGC 1450
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
474 .....ValTyrSerLeuGlyMetThrAsnProTyrPheThrValAsn 487
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1451 GTATGTTTTCGGGAAATTTGAGCTTGACTAATCCAAGGATTTTGGACAGC 1500
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
488 GlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLysTyrAspAs 504
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1501 TGGTATAGCTCTACGATCAATCTTTAT.....GCGGATTA 1535
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
504 nLysAsnIleSerAsnTyrValLeuAspSerTyrGlySerLeuSerT 521
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1536 CAGGATAAGCTACCATATACCAACAGCGGGGGCTTTGGGGTGAATG 1585
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
521 yrGlyTyrProIleAspGluAsnGlnArgIleSerPheGlyLeuAsnAla 537
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1586 TCGGGCGCATGCTGGGTAAATAGAACCCTATGTGAGCTTAGGTATAACTTG 1635
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
538 AspAsnThrLysLeuHisGly 544
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1636 AATGTTACCAAACTCCTTGGT 1656
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

```







```

56  rgLeuGlyGlnValValSerGluAsnGlnLeuAlaAspGlyValLysAla 72
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
214 GTCCGGTCAGCGTGTGACTGACAATGATGCGTAATATGTGCGCTCT 263
73  LeuTyraLathrGlyAsnPhSerAspValGlnValTyHisGlnGlu 89
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
264 TTATTCGTAAGTGTGCGATGCGATGATGTAAGCG...CATCAAGAGG 310
89  YaArgile...IleTyrglnValThrGluArgProLeuIleAlaGluIleA 105
   | : : : : : : : : : : : : : : : : : : : : : : : : :
311 CGATGTCGTTGTTGTTAGCGGTTGGCGTAATCGATCATTTTCAGATGTTA 360
105 snPheGluGlyAsnArgLeuIleProLysGluGlyGlnGluGlyLeu 121
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
361 AAATCAAGGTAACCTCGTTATTCCTCCACTGAAGCAGCTTAAACAAACTTA 410
122 LysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrValG 138
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
411 GATGCTAACGGGTTTAAAGTTGGCGATGTTTAAATTCGAGAAAAATTA 460
138 nMetIleGluThrGluLeuThrAsnGlnTyIleSerGlnGlyTyrrA 155
   | : : : : : : : : : : : : : : : : : : : : : : : : :
461 TGAATTTGCCAAAGGTGTAAGAGCACTATGCAAGGTAGTGGCTATA 510
155 snThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgValLys 171
   ||||| : : : : : : : : : : : : : : : : : : : : : : :
511 ACGCAACAGTTGAACCTATTGTCAATACGCTACCAATAATCGCGCTGAA 560
172 LeuAspMetThrPheAlaGluGlyLysProAlaArgValValAspIleAs 188
   ||||| : : : : : : : : : : : : : : : : : : : : : : :
561 ATTTTAATTCAAATCAATGAAGATGATAAGCAAAATTTGCGCATCAATTA 610
188 nIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspValLeuA 205
   : : : : : ||||| : : : : : : : : : : : : : : : : : :
611 TTTCAGGGGAACGAATCTGTTAGTACAGTACATTTACAAGACAAATGG 660
205 laIleLysAspAsnLysIleAsnProLeuSerLysAlaAspArgTyThr 221
   : : : : : : : : : : : : : : : : : : : : : : : : :
661 AATTACAACCTGATCTCTGTTGGAAATTA...TGGGGAATAAATTTGAA 707
222 GlnGluLysLeuValThrSerLeuGluAsnLeuArgAlaLysTyrrLeuAs 238
   : : : : : : : : : : : : : : : : : : : : : : : : :
708 GGTGCGCAATTCGAGAAAGATTTGCACTCAATTCGTTATTTATTTAA 757
238 nAlaGlyPheValArgPheGluIleLysAspAlaLysLeuAsnIleAsnG 255
   | ||| : : : : : : : : : : : : : : : : : : : : : : :
758 TAATGGCTATGCCAAACACAAATTAATAAACGGATGTTCAAGTAATG 807
255 luAspLysAsnArgIlePheValGluIleSerLeuHisGluGlyGluGln 271
   : : : : : : : : : : : : : : : : : : : : : : : : :
808 ATGAAAAACAAAGTTAATGTAACCATTTGATGTAATGAAGTTTACAG 857
272 TyrrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuThrTyrrGlnAl 288
   ||||| : : : : : : : : : : : : : : : : : : : : : : :
858 TATGACCTTCGTAGTGACGCATTTATAGGTAATCTGGGAGGTATGTCG 907
288 aGluLeuGluAlaLeuLeuLys...PheLysAlaGluGluGlyPheSerG 304
   ||||| ||||| ||||| : : : : : : : : : : : : : : : :
908 CGAGCTTGAACCTTTACTTTACAGCATTTACATTTAAATGATACTTTCGCC 957
304 lnAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAsp 320
   : : : : : : : : : : : : : : : : : : : : : : : : :
958 GTAGTGATATTGCAGATGTAGAAATGCAATTAAGCAAAACTTGGAGAA 1007
321 AspGlyTyrrTyrrAlaGlnIleArgProValThrArgIleAsnAspG 337
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1008 CGCGGTTACGGTAGCGCAACGGTAATTCAGTACCTGATTTTGTATGATGC 1057
337 uSerArgThrValAspValGluTyrrTyrrIleAspProValHisProVal 354
   : : : : : : : : : : : : : : : : : : : : : : : : :
1058 AAATAAACAATAGCGATAACCCCTTGTGTTGATGCTGGACGAGCTTAA 1107

```

```

354 yrValArgArgIleAsnPhThrGlyAsnPhelyThrGlnAspGluVal 370
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1108 CTGTTGCCCAACTCGCTTTGAAGAAATACCGTTCTGCTGATAGCACT 1157
371 LeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLy 387
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1158 TTACGTCAGGAATGCCCAACGAAGAACTTGGTATATATTCACAAT 1207
387 sIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePheLysHisV 404
   : : : : : ||||| : : : : : : : : : : : : : : : : : :
1208 AGTTGAGTTAGAAAAATTCGCTTAGATCGTACAGGTTTCTCGAA... 1253
404 alThrValAspThrArg.....ProValProAsnSerProAspGlnVal 418
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1254 ..ACAGTCGAAAAACCGAATTCATCTATCAATGCTAGTAGTAAGAGTG 1301
419 AspValAsnPheValValGluGluGlnProSerGlySerSerThrIleAl 435
   ||||| : : : : : ||||| : : : : : : : : : : : : : : :
1302 GATGTCGTATATAAAGTCAAGAACGTAACACGGGTAGTATCAACTTTGG 1351
435 aAlaGlyTyrrSerGlnSerGlyGlyValThrPheGlnPheAspValSerG 452
   : ||||| : : : : : ||||| : : : : : : : : : : : : : : :
1352 TATGTTGTCGTTACAGAGAGTGGTATTAGTTATCAAGCAAGTGTAAAC 1401
452 lnAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1402 AAGATAATTTCTTGGGAACAGGGCGGCAGTAGTATAGCTGGTACGAAA 1451
469 SerGluThrArgGluValTyrrSerLeuGlyMetThrAsnProTyrrPheTh 485
   : : : : : : : : : : : : : : : : : : : : : : : : :
1452 AATGATTATGTTGTCAGAGTGTCAATTTGGTGTATACCGAGCCCTATTAT 1501
485 rValAsnGlyValSerGlnSerLeuSerGlyTyrrTyrrArgLysThrLys 502
   : : : : : ||||| : : : : : : : : : : : : : : : : : :
1502 TAAAGATGGTGA.....AGTCTGCTGGCAATGTTTCTTTCAAAACT 1545
502 yrAspAsnLys.....AsnIleSerAsnTyrrValLeuAspSerTyrr 515
   ||||| ||||| : : : ||||| ||||| : : : |||||
1546 ACGATAACTCTAAAAGTGATACATCTCTTAACATATAAGCGTACGACTTAC 1595
516 GlyGlySerLeuSerTyrrGlyTyrrProIleAspGluAsnGlnArgIleSe 532
   ||||| : : : : : ||||| : : : : : : : : : : : : : : :
1596 GGAAGTAATCTTACTTTAGGTTTCCCTGTAAATGAAATAACTCCTATTA 1645
532 rPheGlyLeu.....AsnAlaAspAsnThrLysLeuHisG 544
   ||||| : : : : : : : : : : : : : : : : : : : : : :
1646 TGTAGGATTAGTCACTACCTATATAATAAATAGTAACCTTTGCTCTAGAA 1695
544 lyGlyArgPheMetGlyIleSerAsnValLysGlnLeuMetAlaAspGly 560
   : : : : : : : : : : : : : : : : : : : : : : : : :
1696 ATAACCGTAATTTATATATTTCAATCAATGAAA..... 1727
561 GlyLysIleGlnValAspAsnAsnGlyIleProAspPheLysHisAspTy 577
   : : : : : ||||| : : : : : : : : : : : : : : : : : :
1728 .....TTTAAAGTAATGGCATTAAACAAATGACTTTGATTT 1765
577 rThrThrTyrrAsnAlaIleLeuGlyTyrrAsnTyrrSerSerLeuAspArg 594
   : : : : : ||||| : : : : : : : : : : : : : : : : : :
1766 TTCT.....TTTGGTTGGAACATAACAGCCTTAATAGAG 1800
594 roValPheProThrGlnGlyMetSerHisSerValAsp.....LeuThr 608
   ||||| ||||| ||||| : : : : : : : : : : : : : : : :
1801 GCTATTTCCCAACTAAAGGGGTTAAAGCAAGTCTTGGTGGACGAGTTACT 1850
609 ValGlyPheGlyAspLysThrHisGlnLysValValTyrrGlnGlyAsnI 625
   : : : : : ||||| : : : : : : : : : : : : : : : : : :
1851 ATCCAGGTTCTGATACAAATACTACAAACTAAGTCCAGATGTACAGGG 1900
625 eTyrrArgProPhe.....IleLysLysSerValL 635
   : : : : : ||||| : : : : : : : : : : : : : : : : : :
1901 TTTCTACCCCATAGACAGACATCACCTCTGGGTTGTATCTGCAAAAGCAT 1950
635 euArgGlyTyrrAlaLysLeuGlyTyrrGlyAsn...AsnLeuProPheTyrr 650

```



[illegible]

469 SerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyPheth 485  
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
1452 AATGAATTATGGTCACGAGTGCTCAAAATTTGGGTATATCCGAGCGCCTATTTTAC 1501  
rValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrlYst 502  
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
1502 TAAAGATGGTGTA.....AGTCITGGTGGAATAVTTTTCTTTGAAAAC 1545  
502 TyrAsnLys.....AsnIleSerAsnTrpValLeuAspSerTyr 515  
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
1546 ACGATAACTCTAAAGTAGATACATCCTCTAACAATAGCGTACGACTTAC 1595  
516 GlyGlySerLeuSerTyrGlyTyrProIleaspGluAsnGlnArgIleSe 532  
||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
1596 GGAAGCTAATGTACTTTAGGTTTCCTCTGAATAGAAATAACTCCTATTAT 1645  
532 rPheGlyLeu.....AsnAlaAspAsnThrlYsLeuHisG 544  
|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
1646 TGTAGATTAGGTGCATCATATAATAAATTAGTAACTTGCCTAGAAT 1695  
544 lylGlyArgPheMetGlylleSerAsnVallysGlnLeuMetalaaSpGly 560  
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
1696 ATAACCGTAATTTATATATATCAATCAATGAA..... 1727  
561 GlyLysIleGlnValasPasnasnClytleproaspPheLyshisAspty 577  
:::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
1728 .....TTTAAAGGTAATGGCATAAAAAACAATAGCTTGATTT 1765  
577 rThrTrpTyrAsnAlaIleLeuGlyTrpAsnTyrSerSerLeuAspargp 594  
|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
1766 TTCT.....TTTGGTTGGAACATAAACAGCCTTAATACAG 1800  
594 roValPheProThrclncglyMetSerHisServalasp.....LeuThr 608  
|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
1801 GCTATTTCCCRACTAAGGGGTAAAGCAAGCTTGGTGACGAGTTFACT 1850  
609 ValGlyPheGlyAsPLySThrHisGlnLysValValTyrGlnGlyAsnIl 625  
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
1851 ATTCCAGGTTCTGTGATACAAATACTACAAACTAGTCAGATGTACAGGG 1900  
625 eTryArgprophe.....lleylsysServall 635  
|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
1901 TTCTACCACCTTAGACAGATCACCTCTGGTTGTATCTGCAAAAGCAT 1950  
635 euARGlytyrAlalysLeucgltYrGLyasN...asnLeuPropheTyr 650  
|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
1951 CTGCAGGATATGCAAT...GGTTTGGAAACAAGCGCTTTACCGTCTAT 1997  
651 GluAsnPheTYraGLyGlytYrGlyserValargGlyTyraspGlnse 667  
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
1998 CAACCTTATACGGCGGTGGCATCGGTTCATACGTGGTTTGGTTATGG 2047  
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThralaArgGlycLnG 684  
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
2048 TAGTATTGGACCTTAACGCA.....ATTATGCCGATATGGTAATG 2088  
684 lnThrThrLeuGly.....GluvalValGlyGly 693  
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
2089 GTAGTGCTACTGGTACTTTTAAAGAGATAAGTTCGTGATGTGATTTGGT 2138  
694 AsnAlaLeuAlaThrPheGlySerGluLeuIleLeuproLeuProPheLy 710  
|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
2139 ATGCANATCGGTACAGTACGGCGACAGGTTAATGTGGCCAACCTCATTTGT 2188  
710 selyAsPrPlple...aspGlnValargProValilePheileGlcUcyg 726  
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
2189 GAGCGATAAGAGCAAAATACCGTCCGAACCTCCTATTATTTGTTGATCGGG 2238  
726 lyGlnValPheaspThr....GlyMetaspySGlnThrIleaspLeu 741  
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
2239 CAAGTGTTTGAAGTAACATAATGGAAATACAGATAAAATGATTAGACG 2288  
742 ThrGlnPheLysaspProGlnAlaThrAlaclucInAsnAlalyalaAl 758

```
2289 GATGATTAAAA..... 2300
758 aAsnArgProLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyV 775
2301 .AGATTCCTGATTATGGCAAAATCAAGCCGATTCGGCGCTCTACAGGTG 2349
775 alGlyAlaThrTyrThrProLeuGlyProLeuSerIleSerTyrAla 791
2350 TCGGATTCCAATGCAATCCTTATGGCCATTTGGTATTTCTCTATGCC 2399
792 LysProLeuAsnLysLysGlnAsnAspGlnThrAspThrValGlnPheG 808
2400 AAACCAATTAATAATGAAATGATGATGTCGACAGTTCGAACAGTTCCAATTAG 2449
808 nileGlySerValPhe 813
2450 TATTTGGAGGTTCTTTC 2465

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-942-046-1

seq_documentation_block:
; Sequence 1, Application US/08942046
; Patent No. 6284954
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/942,046
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-732 MIS-jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2949 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 75..2465
; US-08-942-046-1
```

```
alignment_scores:
Quality: 1052.50      Length: 822
Ratio: 1.956          Gaps: 20
Percent Similarity: 65.450 Percent Identity: 31.630

alignment_block:
US-09-701-711-2 x US-08-942-046-1 ..
Align seg 1/1 to: US-08-942-046-1 from: 1 to: 2949

23 SerThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIleThrIleTh 39
114 ACAACGACTGTGTTTGGCCACCTTTTGTGCAAAAGATATTCGTGTGA 163
39 rGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPheA 56
164 TGGTGTTCAGGTGACTTAGAACACAAATCCGAGCAAGTTTACCTGTC 213
56 rgLeuGlyGlnValValserGluAsnGlnLeuAlaAspGlyValLysAla 72
214 GTGCCGGTCAAGCGTGTGACTGACAATGATGTGGCTAATATTGTCGCT 263
73 LeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisGlnGluG 89
264 TTATTCGTAAGTGTGCTGATTCGATGTGAAAGCG...CATCAAGAAG 310
89 yArgIle...ileTyrGlnValThrGluArgProLeuIleAlaGluIleA 105
311 CGATGTGCTGTTGTTAGCGTGTGGCTAAATCGATCATTTTCAGATGTTA 360
105 snPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluGlyLeu 121
361 AAATCAAAAGGTAAGTCTGTTATCCCACTGAAGCACTTTAAACAAACIT 410
122 LysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrValG 138
411 GATGCTAACGGGTTAAAGTTGGCGATGTTTAAATTCGAGAAATTTAA 460
138 nMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyrTyr 155
461 TGAATTTGCCAAAAGTGTAAAGAGCACTATGCAAGTGTAGTGCCTATA 510
155 snThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgValLys 171
511 ACGCAACAGTTGAACCTATTGCTCAATACGCTACCAATAATATCGCGCT 560
172 LeuAspMetThrPheAlaGluGlyLysProAlaArgValValAspIleA 188
561 ATTTTAATTCAAATCAATGAAGATGATAAAGCAAAATTTGGCATCTAAC 610
188 nIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspValLeu 205
611 TTCAAGGGGAACCAATCTGTTAGTAGCAGTACATTTCAAGAACAAATGG 660
205 laIleLysAspAsnLysIleAsnProLeuSerLysAlaAspArgTyrThr 221
661 AATTACAACCTGATTCCTGGTGGGAATTA...TGGGGAAATAAATTTGAA 707
222 GlnGluLysLeuValThrSerLeuGluAsnLeuArgAlaLysTyrLeuAs 238
708 GGTGCGCAATTCGAGAAAGATTTCAGTCAATTCGTGATTTATTTAAA 757
238 nAlaGlyPheValArgPheGluIleLysAspAlaLysLeuAsnIleAsn 255
758 TAATGGCTATGCCAAAGCACAAATTTACTAAACGGATGTTTCAGCTAAAT 807
255 luAspLysAsnArgIlePheValGluIleSerLeuHisGluGlyGluGln 271
808 ATGAAAAACAAAGTTAATGCTAACCATTTGATTAATGAAGGTTTACAG 857
272 TyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAl 288
858 TATGACCTTCGTAGTCGACCGCATTTAGGTAATCTGGGAGGTATGCTGTC 907
```

```

288 agluLeuGluAlaLeuLeuLys...PheLysAlaGluGluGlyPheSerG 304
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
908 CGAGCTTGAACCTTTACTTTTACGATTACATTTAAATGATCTTTCGCC 957
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
304 lnaAlaMetLeuGluGlnThrThrAsnAsnLleSerThrLysPheGlyAsp 320
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
958 GTAGTGATATGCAGATGAGAAAATGCAATTTAAAGCAAACTTGAGAA 1007
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
321 AspGlyTyrTyrAlaGlnLleArgProValThrArgLleAsnAspG1 337
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1008 CCGGTTACGGTAGCCACGCGTAATTCAGTACCTGATTTTGATGATGC 1057
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
337 uSerArgThrValAspValGluTyrTyrLleAspProValHisProValT 354
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1058 AAATAAACATTAGCGATAACCTTGTGTGATGCTGGACGCGTTTAA 1107
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
354 yrValArgArgLleAsnPheThrGlyAsnPhelysthrGlnaspGluVal 370
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1108 CTGTTCCCACTTCGCTTTGAAGGAAATACCGTTTCTGCTGATAGCACT 1157
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
371 LeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLy 387
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1158 TTACGTCAGGAATGCGCCCAACAAGAAGAACTTGGTATAATTCACAAT 1207
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
387 slleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePheLysHisV 404
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1208 AGTTGAGTTAGAAAATTCGCTTAGATCGTACAGGTTTCTTCGAA... 1253
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
404 alThrValAspThrArg.....ProValProAsnSerProAspGlnVal 418
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1254 ..ACAGTGAACCAATGATGCTCAATGCTCAATGCTAGTAAATGAAGTG 1301
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
419 AspValAsnPheValValGluGlnProSerGlySerSerThrIleAl 435
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1302 GATGCTATATAAAGTCAAGAAAGTAAACACGCGTAGTATCAACTTGG 1351
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
435 aalaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSerG 452
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1352 TATGTTACGTCAGAGAGTGGTATTAGTTATCAAGCAAGTGTAAAC 1401
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
452 lnaAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1402 AGATAAATTTCTGGACACGCGCGAGTAGTATAGTACGTCGTACGAAA 1451
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
469 SerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheTh 485
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1452 AATGATTATGTTACGAGTCTCAATTTGGGTTATACCGAGCCCTATTTTAC 1501
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
485 rValAsnGlyValSerGlnSerLeuSerGlyTyrArgLysThrLysT 502
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1502 TAAAGATGGTGA.....AGTCTTGGTGGAATGTTTCTTTTGAAAACT 1545
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
502 yTAspAsnLys.....AsnLleSerAsnTyrValLeuAspSerTyr 515
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1546 ACGATAACTTAAAGTATACATCTCTAATAATAAGCGTACGACTTAC 1595
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
516 GlyGlySerLeuSerTyrGlyTyrProLleAspGluAsnGlnArgLleSe 532
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1596 GGAAGTAATGTACTTTAGTTTCCCTGCTGTAATGAAATAAATCTCTATTA 1645
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
532 rPheGlyLeu.....AsnAlaAspAsnThrLysLeuHisG 544
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1646 TGTAGGATTAGCTACCTATAATAATAATAGTAACCTTGTCTAGAT 1695
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
544 lYgLYArgPheMetGlyLleSerAsnValLysGlnLeuMetaLalaSpGly 560
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1696 ATAACCGTAATTTATATATCAATCAATCAAGAAA..... 1727
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
561 GlyLysLleGlnValAspAsnAsnGlyLleProAspPheLysHisAspTy 577
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1728 .....TTTAAAGGTAATGGCAATTAACAAATGACTTTGATT 1765
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```
seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-433-522A-5
```

```
seq_documentation_block:
```

```
; Sequence 5, Application US/08433522A
```

```
; Patent No. 6013514
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: CHONG, Pele
```

```
; APPLICANT: THOMAS, Wayne
```

```
; APPLICANT: YANG, Yan Ping
```

```

; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433,522A
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-434 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2950 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 334..2724
; US-08-433-522A-5

```

## alignment\_scores:

```

Quality: 1052.50      Length: 822
Ratio: 1.956          Gaps: 20
Percent Similarity: 65.450 Percent Identity: 31.630

```

## alignment\_block:

```
US-09-701-711-2 x US-08-433-522A-5 ..
```

```
Align seg 1/1 to: US-08-433-522A-5 from: 1 to: 2950
```

```

23 SerThrHisAlaGlnAlaAspPheMetAlaAsnAspIleThrIleTh 39
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
373 ACAACGACTGTGTTGCCGCACCTTTTGTGGCAAGATATTTCGTGGA 422
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
39 rGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPheA 56
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
423 TGGTGTTCAGGTGACTTAGACACACAAATCCGACGAGTTTACCTGTC 472
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
56 rgLeuGlyGlnValSerGluAsnGlnLeuAlaAspGlyVallysAla 72
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
473 GTGCCGGTCAGCGTGTGCTGACGATGATGCTGCTAAATATTGCGCTCT 522
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
73 LeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisGlnGlu 89
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
523 TTATTCGTAAGTGGTGGATTCGATGATGTAAGGCG...CATCAAGAAGG 569
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
89 yArgile...IleTyrGlnValThrGluArgProLeuIleAlaGluIleA 105
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
570 CGATGTGCTGTGTAGCGTTGTGCTAAATCGATCATTTTCAGATGTTA 619
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
105 snPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGlyLeu 121
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

```

620 AAATCAAAGGTAACTCTGTATTATCCCACTGAAGCACTTAAACAAAACTTA 669
122 LysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrValGI 138
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
670 GATGCTAACGGGTTTAAAGTTGGCGATGTTTAAATTCGAGAAAAATATA 719
138 nMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyrTyrA 155
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
720 TGAATTTGCCAAAAGGTAAAGAGACACATATGCAAGTGTAGTGCCTATA 769
155 snThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgValLys 171
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
770 AGCAACAGTTGAACCTATTGTCAATACGCTACCAATAATTCGCTGAA 819
172 LeuAspMetThrPheAlaGlyLysProAlaArgValValAspIleAs 188
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
820 ATTTTAAATTCAAATCAATGAAGATGATAAGCAAAATTCGCATCATTA 869
188 nIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspValLeuA 205
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
870 TTTCAGGGGGAACGAATCTGTTAGTAGCAGTACATTACAAGACAATGG 919
205 laIleLysAspAsnLysIleAsnProLeuSerLysAlaAspArgTyrThr 221
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
920 AATTACACACTGTATCTTGGTGGAAATTA...TGGGGAATAAATTTGAA 966
222 GlnGluLysLeuValThrSerLeuGluAsnLeuArgAlaLysTyrLeuAs 238
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
967 GTGCGCAATTCGAGAAAGATTTGCAGTCAATTCGTGATTATTATTATA 1016
238 nAlaGlyPheValArgPheGluIleLysAspAlaLysLeuAsnIleAsnG 255
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1017 TAATGCTATGCCAAAGCACAAATTAATAACGGGATGTTCACTAAATG 1066
255 luAspLysAsnArgIlePheValGluIleSerLeuHisGluGlyGluGln 271
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1067 ATGAAAAACAAAAAGTTAATGTAACCATTTGATGATAAATGAAGTTTAC 1116
272 TyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAl 288
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1117 TATGACCTTCGTAGTCACGCATTTATAGGTAATCTGGGAGGTATGCTGC 1166
288 aGluLeuGluAlaLeuLeuLys...PheLysAlaGluGluGlyPheSerG 304
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1167 CGAGCTTGAACTTTACTTTCAGCATTACATTAATGATACTTTCGCC 1216
304 InAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAsp 320
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1217 GTAGTGATATTGCAGATGTAGAAATGCAATTAAGCAAACTTGGAGAA 1266
321 AspGlyTyrTyrTyrAlaGlnIleArgProValThrArgIleAsnAspGI 337
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1267 CGCGGTACGGTAGCGCAACGGTAAATTCAGTACCTGATTTTGATGATGC 1316
337 uSerArgThrValAspValGluTyrTyrIleAspProValHisProValT 354
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1317 AAATAAACAATTAGCATTAACCTTTGTTGATGCTGGCAGCAGCTTAA 1366
354 yrValArgIleAsnPheThrGlyAsnPheLysThrGlnAspGluVal 370
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1367 CTGTTCCCAACTTCGCTTTGAAGGAAATACCGTTTCTGCTGATAGCACT 1416
371 LeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLy 387
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1417 TTACGTCAGGAATGCGCAACAAGAACTTGGTATAAATTCACAATT 1466
387 stIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePheLysHisV 404
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1467 AGTTGATTAGGAAAAATTCGCTTAGTACGACAGGTTTCTTCGAA... 1512
404 alThrValAspThrArg.....ProValProAsnSerProAspGlnVal 418
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1513 ..ACAGTCGAAAAACCAATTTGATCCTTATCAATGTTAGTAGTAATGATGA 1560

```

```
419 AspValAsnPhValValGluGluGlnProSerGlySerSerThrIleAl 435
||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1561 GATGTCGTATATAAGTCAAGAAGCGTAACACGGGTAGTATCAACTTGG 1610
435 aAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSerG 452
|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1611 TATTGGTTACGGTACAGAGTGGTATTAGTTATATCAAGCAAGTGTAAAC 1660
452 InAsnAsnPhMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468
|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1661 AAGATAATTTCTTGGGAACAGGGCGGCGAGTAAGTATAGCTGGTACGAAA 1710
469 SerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheTh 485
||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1711 AATGATTATGTACGAGTGTCAATTTGGGTTATACCGAGCCCTATTTTAC 1760
485 rValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLysT 502
|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1761 TAAAGATGGTGTATATATTTGGGAAATGTTTCTTTGAAAACT 1804
502 YrAspAsnLys.....AsnIleSerAsnTyrValLeuAspSerTyr 515
||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1805 ACGATAACTCTAAAGTCATACATCTCTAATCTATTAACGCTACGACTTAC 1854
516 GlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSe 532
||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1855 GGAAGTAAATGTACTTTAGGTTTCCCTGTAAATGAAATAACTCTCTATTA 1904
532 rPheGlyLeu.....AsnAlaAspAsnThrLysLeuHisG 544
||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1905 TGTAGGATTAGCTCATACCTATATAATAATTAGTAACCTTCTCTAGAA 1954
544 lyGlyArgPheMetGlyLysSerAsnValLysGlnLeuMetAlaAspGly 560
||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1955 ATAACCGTAATTTATATATCAATCAATGAAA..... 1986
561 GlyLysIleGlnValAspAsnAsnGlyIleProAspPheLysHisAspTy 577
||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1987 .....TTTAAAGGTAATGGCATTAATAAAACAAATGACTTTGATTT 2024
577 rThrThrTyrAsnAlaIleLeuGlyTyrAsnTyrSerSerLeuAspArgp 594
||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2025 TTCT.....TTTGGTGGAACTATAACAGCTTAAATAGAG 2059
594 roValPheProThrGlnGlyMetSerHisSerValasp.....LeuThr 608
||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2060 GCTATTTCCCAACTAAAGGGGTTAAAGCAAGTCTTGGTGGACGAGTTACT 2109
609 ValGlyPheGlyAspLysThrHisGlnLysValValTyrGlnGlyAsnII 625
|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2110 ATCCAGGTTCTGATAACAATACTACAATAAGTGCAGATGTACAGGG 2159
625 eTyrArgProPhe.....IleLysLysSerValL 635
|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2160 TTTCTACCCATTACAGAGATCACTCTGGTGTATCTGCAAAAGCAT 2209
635 euArgGlyTyrAlaLysLeuGlyTyrGlyAsn...AsnLeuProPheTyr 650
||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2210 CTGCAGGATATGCAAAAT...GGTTTGGAAACAAGCGTTTACCCTTCTAT 2256
651 GluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667
||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2257 CAAACTTATACAGCGGTGGCATCGGTTTCTATAGCTGGTTTGGTTATGG 2306
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2307 TAGTATGGACCTAACCA.....ATTATGCCGAATATGTAATG 2347
684 InThrThrLeuGly.....GluValValGlyGly 693
|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
2348 GTAGTGCTAGTGGTACTTTTAAAGAAGATAAGTCTGTATGTGTTGGTGT 2397
```

seq\_name: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:US-09-135-166-5

seq documentation block:

; Sequence 5, Application US/09135166

; Patent No. 6083743

; GENERAL INFORMATION:

; APPLICANT: CHONG, Pele

; APPLICANT: THOMAS, Wayne

; APPLICANT: YANG, Yan Ping

; APPLICANT: LOOMORE, Sheena

; APPLICANT: SIA, Dwo Yuan Charles

; APPLICANT: KLEIN, Michel

; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN

; NUMBER OF SEQUENCES: 55

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sim & McBurney

; STREET: 6TH Floor, 330 University Avenue

; CITY: Toronto

; STATE: Ontario

; COUNTRY: Canada

; ZIP: M5G 1R7

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/135,166

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/433,522

; FILING DATE: 12-SEP-1995

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: STEWART, Michael I

; REGISTRATION NUMBER: 24,973

; REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (416) 595-1155

; TELEFAX: (416) 595-1163



```

; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2950 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 334..2724
US-09-135-166-5

```

```

alignment_scores:
  Quality: 1052.50      Length: 822
  Ratio: 1.956          Gaps: 20
  Percent Similarity: 65.450  Percent Identity: 31.630

alignment_block:
US-09-701-711-2 x US-09-135-166-5 ..

Align seg 1/1 to: US-09-135-166-5 from: 1 to: 2950

23 SerThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIleThrIleTh 39
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::
373 ACAACGACTGTGTTGCCGCACCTTTTGGCAAAAGATATTCGTGTGGA 422
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::

39 rGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProphea 56
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::
423 TGGTGTTCAGAGTGACTTAGAACACAAATCCGACGCAAGTTTACCTGTT 472
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::

56 rGlyLeuGlnValValSerGluAsnGlnLeuAlaAspGlyValLysAla 72
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::
473 GTGCCGTCAGCGTGTGACTGACATGATGTGGCTAATATTGTCGCTCT 522
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::

73 LeuTyAlaThrGlyAsnPheSerAspValGlnValThrHisGlnGluI 89
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::
523 TTATTTCGTAAGTGTGATTCGATGATGTAAGCG...CATCAAGAAG 569
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::

89 yArgIle...IleTyGlnValThrGluArgProLeuIleAlaGluIleA 105
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::
570 CGATGCTGTTGTTAGCGTGTGGCTAAATCGATCATTTTCAGATGTTA 619
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::

105 snPheGlyAsnArgLeuIleProLysGluGlyLeuGlnGlyLeu 121
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::
620 AAATCAAGSTAACTCTGTTATTCCTCCAGACACTTAACAAACTTA 669
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::

122 LysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrVal 138
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::
670 GATGCTAACGGGTTTAAAGTTGGCGATGTTTAAATTCGAGAAAAATTA 719
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::

138 nMetIleGluThrGluLeuThrAsnGlnTyIleSerGlnGlyTyTyra 155
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::
720 TCAATTTGCCAAAGTGTAAAGAGACACATATCCAGTGTAGTTCGCTATA 769
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::

155 snThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgValLys 171
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::
770 ACGCAACAGTTGAACCTATTGTCAATACGCTACCAATAATCGCGCTGAA 819
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::

172 LeuAspMetThrPheAlaGlyLysProAlaArgValValAspIleAs 188
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::
820 ATTTTAATTCAAATCAATGAAGATGATAAGCAAAATTCGCATCAATTA 869
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::

188 nIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspValLeu 205
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::
870 TTTCAAGGGGACGAACTGTTAGTACGAGTACATTACAGCAACAATGG 919
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::

205 laileLysAspAsnLysIleAsnProLeuSerLysAlaAspArgTyThr 221
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::
920 AATTACAACCTGATCTTGGTGGAAATTA...TGGGAAATAAATTTGAA 966
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::

222 GlnGluLysLeuValThrSerLeuGluAsnLeuArgAlaLysTyLeuAs 238
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::

```

```

967 GGTGCGCAATTCGAGAAAGATTTCAGTCAATTCGTGATTATTATTTAAA 1016
238 nAlaGlyPheValArgPheGluIleLysAspAlaLysLeuAsnIleAsnG 255
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::
1017 TAATGGCTATGCCAAACACAAATTAATAAACGGATGTTTCAGCTAAATG 1066
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::

255 luAspLysAsnArgIlePheValGluIleSerLeuHisGluGlyGluGln 271
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::
1067 ATGAAAAAACAAAAGTTAATGTAAACCATTTGATGTAATGAAGGTTTAC 1116
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::

272 TyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuThrTyThrGlnAl 288
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::
1117 TATGACCTTCGTAGTCACGCATTATAGTAATCTGGGAGGTATGCTGCG 1166
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::

288 aGluLeuGluAlaLeuLys...PheLysAlaGluGluGlyPheSerG 304
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::
1167 CGAGCTTGAACCTTTACTTTTCAGCATTTACATTAATGATACCTTTCCGCC 1216
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::

304 lnAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAsp 320
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::
1217 GTAGTGTATTCGAGATGTAGAAAAATGCAATTAACGCAAACTTGGAGAA 1266
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::

321 AspGlyTyThrTyAlaGlnIleArgProValThrArgIleAsnAspGl 337
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::
1267 CGCGGTTACGGTAGCGCAACGGTAAATTCAGTACCTGATTTTGATGATGC 1316
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::

337 uSerArgThrValAspValGluTyThrIleAspProValHisProValT 354
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::
1317 AAATAAACATTTAGCGATTAACCTTGTGTTGATGCTGGACGACGTTTAA 1366
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::

354 yrValArgIleAsnPheThrGlyAsnPheLysThrGlnAspGluVal 370
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::
1367 CTGTTCCGCCAACTTCGCTTTGAAGGAATACCGCTTTCGTGCTGATAGCACT 1416
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::

371 LeuArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLys 387
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::
1417 TTACGTGAGAAATGCGCCAAAGAAAGAACTTGGTATTAATTCACAATT 1466
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::

387 sIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePheLysHis 404
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::
1467 AGTTGAGTTAGGAAAAATTCGCTTAGCTACAGGTTCTTCGAA... 1512
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::

404 alThrValAspThrArg.....ProValProAsnSerProAspGlnVal 418
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::
1513 ..ACAGTCGAAACCCGAATTTGATCCTCAATGCTAGTAAATGATGAAGTG 1560
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::

419 AspValAsnPheValValGluGluGlnProSerGlySerSerThrIleAl 435
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::
1561 GATGCTGATATAAAGTCAAAAGAACGTAACACGGGTAGTATCAACTTTGG 1610
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::

435 aAlaGlyTyThrSerGlnSerGlyGlyValThrPheGlnPheAspValSer 452
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::
1611 TATTGTTACGGTACAGAGAGTGTATTAGTTATCAAGCAAGTGTTAAC 1660
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::

452 lnAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::
1661 AAGATAAATTTCTTGGCAACAGGGCGGCAGTAAGTATAGCTGGTACGAAA 1710
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::

469 SerGluThrArgGluValTyThrSerLeuGlyMetThrAsnProTyThr 485
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::
1711 AATGATTAATGTCACGAGTGTCAATTTGGGTTATACCGAGCCCTATTATTAC 1760
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::

485 rValAsnGlyValSerGlnSerLeuSerGlyTyThrTyArgLysThrLys 502
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::
1761 TAAAGATGTTGTA.....AGTCTTGGTGGAAATGTTTCTTTCGAAAAC 1804
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::

502 yrAspAsnLys.....AsnIleSerAsnTyThrValLeuAspSerTy 515
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::
1805 ACGATAACTCTAAAGATGATACATCTCTCAACTATAAGCGTACGACTTAC 1854
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::

516 GlyGlySerLeuSerTyThrGlyTyThrProIleAspGluAsnGlnArg 532
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::
1855 GGAAGTAATGTTACTTTAGTTCCTCCCTGTAATGAAATAAATAACTCCTATTA 1904
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  ::

```

532 rPheGlyLeu.....AsnAlaAspAsnThrLysLeuHisG 544  
||||| : : : : :  
1905 TGTAGGATTAGGTATACCTATAATAAATAGTAACCTTCTCTAGAAT 1954  
: : : : :  
544 lYgIyArgPheMetGlyIleSerAsnValLysGlnLeuMetAlaAspGly 560  
: : : : :  
1955 ATAACCGTAATTATATATCAATCAATGAAA..... 1986  
: : : : :  
561 GlyLysIleGlnValAspAsnGlyIleProAspPheLysHisAspTy 577  
: : : : :  
1987 .....TTTAAAGGTAATGGCAATAAACAAAGACTTTGATT 2024  
: : : : :  
577 rThrThrTyraAsnAlaIleLeuGlyTrpAsnTySerSerLeuAspArgp 594  
: : : : :  
2025 TTCT.....TTTGGTGAAGTATAACAGCCTTATAGAG 2059  
: : : : :  
594 roValPheProThrGlnGlyMetSerHisSerValAsp.....LeuThr 608  
: : : : :  
2060 GCTATTTCCTCAACTAAAGGGGTTAAAGCAAGCTTGTGGAGGAGTTACT 2109  
: : : : :  
609 ValGlyPheGlyAspLysThrHisGlnLysValValTyGlnGlyAsnII 625  
: : : : :  
2110 ATTCCAGGTTCTGATAACAATACTACAACACTTAAGTCAGATGTACAGGG 2159  
: : : : :  
625 eTyArgProPhe.....IleLysLysSerValL 635  
: : : : :  
2160 TTCTACCATATAGACAGATCACCTCTGGGTTGTATCTGCAAAAGCAT 2209  
: : : : :  
635 euArgGlyTyraLysLeuGlyTyGlyAsn....AsnLeuProPheTy 650  
: : : : :  
2210 CTGCAGGATATGCAAAAT..GGTTTGGAAACAAGCGTTTACCGTTCTAT 2256  
: : : : :  
651 GluAsnPheTyraLysGlyTyGlySerValArgGlyTyraSpGlnse 667  
: : : : :  
2257 CAAACTTATACAGCGGGTGGCATCGGTTTACCTATTCAGTTTTCCTATGG 2306  
: : : : :  
667 rSerLeuGlyProArgSerGlnAlaTyrrLeuThrAlaArgArgGlyGlnG 684  
: : : : :  
2307 TAGTATTGACCTAAGCA.....ATTATGCCGAATATGTTAATG 2347  
: : : : :  
684 lInThrThrLeuGly.....GluValValGlyGly 693  
: : : : :  
2348 GTAGTGGTACTGGTACTTTTAAAGAGATAAGTTCTGATGTGATGGTGT 2397  
: : : : :  
694 AsnAlaLeuAlaThrPheGlySerGluLeuIleLeuProLeuProPheTy 710  
: : : : :  
2398 ATGCANTCGCTACAGCTAGCGCAGAGTTAATGTGCCAACTCCATTGT 2447  
: : : : :  
710 sGlyAspTrpIle...AspGlnValArgProValIlePheIleGluGlyG 726  
: : : : :  
2448 GAGCGATAAGAGCAAAATACGTCGGAACCTCCTTATTGTTGATGCGG 2497  
: : : : :  
726 lYcInValPheAspThrThr...GlyMetAspLysGlnThrIleAspLeu 741  
: : : : :  
2498 CAAGTGTGTTGGAACTAATGAATCAGATAAAATGATGATTAGAGACC 2547  
: : : : :  
742 ThrGlnPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAl 758  
: : : : :  
2548 GATGTATTAA..... 2559  
: : : : :  
758 aAsnArgProLeuLeuThrGlnAspLysGlnLeuArgTyrrSerAlaGlyV 775  
: : : : :  
2560 .AGATTGCTGTATTATGGCAAAATCAAGCGGTATTCGCGCTCTACAGGTG 2608  
: : : : :  
775 alGlyAlaThrTrpTyrrThrProIleGlyProLeuSerIleSerTyraAla 791  
: : : : :  
2609 TCGGATTCCAATGGCAATCTCTATTGGGCCATTGTTATCTCTTATGCC 2658  
: : : : :  
792 LysProLeuAsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGl 808  
: : : : :  
2659 AAACCAATTAATAAATATGAATGATGATGTGCGAAGTTCCTCAATTTAG 2708  
: : : : :  
808 nIleGlySerValPhe 813  
: : : : :  
2709 TATTGGAGGTTCTTTC 2724

seq\_name: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:US-08-942-046-5

seq\_documentation\_block:  
; Sequence 5, Application US/08942046  
; Patent No. 6264954  
; GENERAL INFORMATION:  
; APPLICANT: CHONG, Pele  
; APPLICANT: THOMAS, Wayne  
; APPLICANT: YANG, Yan Ping  
; APPLICANT: LOOSMORE, Sheena  
; APPLICANT: SIA, Dwo Yuan Charles  
; APPLICANT: KLEIN, Michel  
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN  
; NUMBER OF SEQUENCES: 55  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6TH Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA: US/08/942,046  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/433,522  
; FILING DATE: 12-SEP-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: STEWART, Michael I  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-732 MIS:jb  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2950 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 334..2724  
; US-08-942-046-5

alignment\_scores:  
Quality: 1052.50 Length: 822  
Ratio: 1.956 Gaps: 20  
Percent Similarity: 65.450 Percent Identity: 31.630  
alignment\_block:  
US-09-701-711-2 x US-08-942-046-5 ..  
Align seg 1/1 to: US-08-942-046-5 from: 1 to: 2950  
23 SerThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIleThrIleTh 39  
: : : : :  
373 ACAACCACTGTGTTTCCCGCACCTTTTGGCAAAAGATATTCTGTGGA 422  
: : : : :  
39 rGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProphea 56



```
635 euArgGlyTyrAlaLysLeuGlyTyrGlyAsn...AsnLeuProPheTyr 650
|||||
2210 CTGCAGGATATGCAAA...GGTTTGGAAACAAGCGTTTACCGTTCTAT 2256
|||||
651 GluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667
|||||
2257 CAACCTATACAGCGGGGGGATCGGTTTCATTAGCTGGTTTGTCTATGG 2306
|||||
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgGlyGlnG 684
|||||
2307 TAGTATTGACCTACGCA...ATTATGCCGAATGATGTAATG 2347
|||||
684 lnThrThrLeuGly...GluValValClyCly 693
|||||
2348 GTAGTGGTACTGGTACTTTTAAAGATAAGTTCTGTGATGTTGGTGGT 2397
|||||
694 AsnAlaLeuAlaThrPheGlySerGluLeuIleLeuProLeuPheLy 710
|||||
2398 AATGCAATCGTACAGTAGCGCAGAGTTAATTGTGCCAACTCCATTGT 2447
|||||
710 sGlyAspTrpIle...AspGlnValArgProValIlePheIleGluGly 726
|||||
2448 GAGCGATAAGACCAAAATACGGTCCGAACTCCTTATTGTTGATCGG 2497
|||||
726 lyGlnValPheAspThrThr...GlyMetAspLysGlnThrIleAspLeu 741
|||||
2498 CAAGTGTGTGGAACTACTAAATGGAAATCAGATAAAATGGATTAGAGAGC 2547
|||||
742 ThrGlnPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLysAla 758
|||||
2548 GATGTTATTAATA... 2559
|||||
758 aAsnArgProLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyV 775
|||||
2560 .AGATTGCTGCTATTATGCAAAATCAAGCGGTATTGCGGCCCTCAGAGTG 2608
|||||
775 alGlyAlaThrTrpThrThrProIleGlyProLeuSerIleSerTyrAla 791
|||||
2609 TCGGATTCATGCAATGCAATCTCTATTGGCCATTGGTATTCTTATGCC 2658
|||||
792 LysProLeuAsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGl 808
|||||
2659 AAACCAATTAATAAATATGAAATGATGATGTCGAACAGTTCCAATTAG 2708
|||||
808 nIleGlySerValPhe 813
|||||
2709 TATTGGAGGTTCTTTC 2724
|||||
```

seq\_name: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:US-08-433-522A-3

```
seq_documentation_block:
; Sequence 3, Application US/08433522A
; Patent No. 6013514
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433,522A
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-434 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2984 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 374..2764
; US-08-433-522A-3
```

```
alignment_scores:
Quality: 1052.50 Length: 822
Ratio: 1.956 Gaps: 20
Percent Similarity: 65.450 Percent Identity: 31.630

alignment_block:
US-09-701-711-2 x US-08-433-522A-3 ..
```

```
Align seg 1/1 to: US-08-433-522A-3 from: 1 to: 2984

23 SerThrHisAlaGlnAlaAspPheMetAlaAsnAspIleThrIleTh 39
413 ACAAGGACTGTGTTGCGCGACCTTTTGTGGCAAAAGATATTGCGTGGGA 462
39 rGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPheA 56
463 TGGTGTTCAGAGTGACTTAGAACACAAATCCGACGAGTTTACCCTGTC 512
56 rGlyLeuGlyValValSerGluAsnGlnLeuAlaAspGlyValLysAla 72
513 GTGCGGTCAGCGTGTGACTGACAAATGATGGCTAATATTGTCGCTCT 562
73 LeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisGlnGluGl 89
563 TTATTCGTAAAGTGTGATTCGATGATGAAAGCG...CATCAAGAAGG 609
89 yArgIle...IleTyrGlnValThrGluArgProLeuIleAlaGluIleA 105
610 CGATGTGCTTGTGTTGCGGTTGGCTAAATTCGATCATTTTCAGATGTTA 659
105 snPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluGlyLeu 121
660 AAATCAAAAGGTAACCTCTGTTATTCCTCCACTGAAGCAGCTTAAACAAA 709
122 LysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrValGl 138
710 GATGCTAACGGGTTTAAAGTTGGCGATGTTTAAATTCGAGAAAAATAAA 759
138 nMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyrTyrA 155
760 TGAATTTGCCAAAAGTGTAAAGACACTATGCAAGTGTAGTTCGCTATA 809
155 snThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgValLys 171
810 ACGCAACAGTTGACCTATTGTCAATAGCTACCAAAATAATCGCGTGA 859
172 LeuAspMetThrPheAlaGluGlyLysProAlaArgValValAspIleAs 188
```

860 ATTTTAATCAATCAATGAAGATGATAAAGCAAAATTTGCCATCAATTAAC 909  
188 nIlelleGlyAsnGlnHisPheSeraspAlaaspLeulleAspValleuA 205  
910 TTTCAAGGGAGCAAGCAATCTGTTAGTAGCAGATACATTACAAGAACAATGG 959  
205 laileLysaspAsnLysIleAsnProLeuSerLysAlaaspArgTyrThr 221  
960 AATTACAACTGATTCTTGCTGGAATTA...TGGGAAATAAATTTGAA 1006  
222 GlnGluLysLeuValThrSerLeuGluAsnLeuArgAlaLysTyrLeuAs 238  
1007 GGTGCGCAATTTCAGAGAAGATTTGCAGTCAATTCGTGATTATTATTA 1056  
238 nAlaGlyPheValArgPheGluIleLysaspAlaLysLeuAsnIleAsnG 255  
1057 TAATGCTATGCCAAAGCAAAATTTACTAAACGGATGTTTCAGCTAAATG 1106  
255 luAspLysAsnArgIlePheValGluLysSerLeuHisGluGlyGluGln 271  
1107 ATCAAAAAACAAAAGTTAATGTAAACCATTCATCTAAATGAAGCTTTACAG 1156  
272 TyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAl 288  
1157 TATGACCTTCGTAGTCAGCAGCATATAGGTAATCTGGAGGTATGCTGC 1206  
288 aGluLeuGluAlaLeuLeuLys...PheLysAlaGluGluGlyPheSerg 304  
1207 CGAGCTTGAACTTACTTTCACGATACATTTAAATGATACTTTCGCC 1256  
304 InAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAsp 320  
1257 GTAGTGATATTCAGATGTAGAAATGCAATTAAGCAAACTTGAGAA 1306  
321 AspGlyTyrTyrThrAlaGlnIleArgProValThrArgIleAsnAspGl 337  
1307 CGCGGTACGGTACGCCACACGGTAAATTCAGTACCTGATTTTGATGTC 1356  
337 uSerArgThrValAspValGluLysTyrIleAspProValHisProValT 354  
1357 AAATAAACCATTAAGCATACCTGTTGTGATGCTGGACGAGCTTAA 1406  
354 yrValArgArgIleAsnPheThrGlyAsnPheLysThrGlnAspGluVal 370  
1407 CTGTCGCCAATCTCGCTTGAAGAAATACCGTTCTGCTGATAGCACT 1456  
371 LeuArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLy 387  
1457 TTACGTCAGGAAATGCCCAACAGAGGAACCTTGGTATATTAATTCACAAT 1506  
387 sIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePheLysHisV 404  
1507 AGTTGAGTTAGGAAATTCGCTAGATGCTAGAGTTCTTCGAA... 1552  
404 alThrValAspThrArg.....ProValProAsnSerProAspGlnVal 418  
1553 ..ACAGTCGAAACCCAAATGATCTTATCAATCGTAGTAATGATGAAGTG 1600  
419 AspValAsnPheValValGluGluGlnProSerGlySerSerThrIleAl 435  
1601 GATGTCGTATATAAGTCAAGAAACGTAACACGGGTAGTATCAACTTTGG 1650  
435 aAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSerg 452  
1651 TATTGTTAGGTACAGAGAGTGGTATTAGTTATCAAGCAAGTGTAAAC 1700  
452 InAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468  
1701 AAGATAAATTTCTTGGGAACAGGGCGGACAGTAAGTAGCTGTCAGCAA 1750  
469 SerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheTh 485

1751 AATGATTATGCTAGAGTGTCAATTTGGGTATATACCGAGCCCTATTTTAC 1800  
485 rValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLysT 502  
1801 TAAAGATGGTGA...AGTCTTGGTGGAATGTTTTCTTTGAAAACT 1844  
502 yrAspAsnLys.....AsnIleSerAsnTyrValLeuAspSerTyr 515  
1845 ACGATACTCTAAAGAGTACATCTCTAATCTAATAGCGTACGACTTAC 1894  
516 GlyClySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSe 532  
1895 GGAAGTAATCTTACTTTTAGTTCCTCTGTAATCAAAATAACTCCTATTA 1944  
532 rPheGlyLeu.....AsnAlaAspAsnThrLysLeuHisG 544  
1945 TGTAGGATAGGTCATACCTATATAATAAATAGTAACCTTGCCTAGAA 1994  
544 lyGlyArgPheMetGlyIleSerAsnValLysGlnLeuMetAlaAspGly 560  
1995 ATAACCGTAATTTATATATATCAATCAATGAAA..... 2026  
561 GlyLysIleGlnValaspAsnAsnGlyIleProaspPheLysHisAspTy 577  
2027 .....TTTAAAGGTAATGCGATTTAAACAAATGACTTTGATTT 2064  
577 rThrThrTyrAsnAlaIleLeuGlyTyrAsnTyrSerSerLeuAspArgp 594  
2065 TTCT.....TTGGTGGAACTATACAGCCTTAAATAGAG 2099  
594 roValPheProThrGlnGlyMetSerHisSerValasp.....LeuThr 608  
2100 GCTATTTCCCACTAAAGGGTTAAACCAAGTCTTGGTGACGAGTTACT 2149  
609 ValGlyPheGlyaspLysThrHisGlnLysValValTyrGlnGlyAsnII 625  
2150 ATTCCAGGTTCTGATACAAATATCTACAACTAAGTCAGATGTACAGGG 2199  
625 eTyrArgProPhe.....IleLysLysSerValL 635  
2200 TTTCTACCCATTAGACAGAGATCACCTCTGGTGTATCTGCAAAAGCAT 2249  
635 euArgGlyTyrAlaLysLeuGlyTyrGlyAsn...AsnLeuProPheTyr 650  
2250 CTCAGGATATGCAAT...GGTTTTGAAACCAAGCGTTTACCGTTCTAT 2296  
651 GluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667  
2297 CAACTTATACAGCGGTGCGATCGGTTTCATTACGTGTTTTTGGCTTATGG 2346  
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684  
2347 TAGTATGGACCTAACGCA.....ATTATGCCGAATATGCTAATG 2387  
684 InThrThrLeuGly.....GluValValGlyGly 693  
2388 GTAGTGTCTGCTGCTTTTAAAGAAGATAAGTCTGATGTGATGGTGGT 2437  
694 AsnAlaLeuAlaThrPheGlySerGluLeuIleLeuProLeuProPheLy 710  
2438 AATGCAATCGCTACAGCTAGCGAGAGTAAATTTGTGCCAACTCCATTTGT 2487  
710 sGlyAspTrpIle...AspGlnValArgProValIlePheIleGluGlyG 726  
2488 GAGCGATAGAGCCAAATACGGTCCGAACCTCCTTATTGTTGATCGGG 2537  
726 lyGlnValPheAspThrThr...GlyMetAspLysGlnThrIleAspLeu 741  
2538 CAAGTGTTCGGAATCTAATGGAATACAGATAAATGATGATTAGAGAGC 2587  
742 ThrGlnPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaI 758  
2588 GATGTATTAAAA..... 2599

```

758 aAsnArgProLeuLeuThrGlnAspLysGlnLeuAurTyrSerAlaGlyV 775
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2600 .AGATTGCTGATTATTGGCAATCAAGCGGTATTTCGGCGCTCTACAGGTG 2648
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

775 aGlyAlaThrTrpTyrThrProIleGlyProLeuSerIleSerTyrAla 791
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2649 TCGGATTCCAATGGCAATCTCCTATTGGCGCATTTGGTATTCTCTATGGC 2698
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

792 LysProLeuAsnLysLysGlnAsnAspGlnThrAspThrValGlnPheG1 808
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2699 AAACCAATAAAAAATGAAAATGATGATGATGTCGAAACAGTTCCAATTAG 2748
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

808 nIleGlySerValPhe 813
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2749 TATTGGAGGTCTTTTC 2764

```

```

Ratio: 1.956      Caps: 20
Percent Similarity: 65.450      Percent Identity: 31.630

alignment_block:
US-09-701-711-2 x US-09-135-166-3  ..

Align seg 1/1  to: US-09-135-166-3  from: 1  to: 2984

23  SerThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIleThrIle 39
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
413 ACAACGACGTGCTTTGCCGACACTTTTGTGGCAAAAGATATTTCGTGTGGA 462

39  rClyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPhe 56
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
463 TGGTGTTCAGGTGACTTTAGAACAAATCCGAGCAAGTTTACCTGTTTC 512

56  rLeuGlyGlnValValSerGluAsnGlnLeuAlaAspGlyVallysAla 72
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
513 GTGCCGGTCAGCGTGACTGACATGATGGCTAAATATTGTCGCTCT 562

73  LeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisGlnGlu 89
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
563 TTATTTCGTAAGTGGTGATTCGATGATGGAAGCG...CATCAAGAAG 609

89  yArgile...IleTyrGlnValThrGluArgProLeuIleAlaGluIle 105
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
610 CGATGTGCTGTGTGTAGCGTTGTGGCTAAATCGCATCAATTCAGATGTTA 659

105  snPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluGlyLeu 121
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
660 AAATCAAAAGGTAACTCTGTATTCCCACTGAAGCACTTAAACAAACTTA 709

122  LysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrValG 138
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
710 GATGCTAACGGGTAAAGTTGGCGATGTTTAATTCGAGAAAATATA 759

138  nMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyrTy 155
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
760 TGAATTTGCCAAAAGTGTAAAAGAGACACTATGCAAGTAGTAGTTCGTATA 809

155  snThrClnIleThrValLysGlnThrMetLeuAspGlyAsnArgValLys 171
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
810 ACGACACAGCTGAACCTATTGTCAATACGCTACCAATAATATCGCGTGAA 859

172  LeuAspMetThrPheAlaGluGlyLysProAlaArgValValAspIle 188
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
860 ATTTTAATTCAAATCAATGAAGATGATAAGCAAAATTTGCGATCATTAAC 909

188  nIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspValLeu 205
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
910 TTTCAAGGGGAACGAATCTGTAGTAGCAGTACATTACAGAACAATGG 959

205  laIleLysAspAsnLysIleAsnProLeuSerLysAlaAspArgTyrThr 221
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
960 AATTACAACCTGATCTTGGTGAAATTA...TGGGGAATAATAATTGAA 1006

222  GlnGluLysLeuValThrSerLeuGluAsnLeuArgAlaLysTyrLeu 238
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1007 GGTGCGCAATTCGAGAAAGATTTCAGTCAATTCGTATTATTATTTAAA 1056

238  nAlaGlyPheValArgPheGluIleLysAspAlaLysLeuAsnIleAsn 255
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1057 TAATGGCTATGCCAAGCACAAATTTACTAAACGGATGTTCACTTAATG 1106

255  luAspLysAsnArgIlePheValGluIleSerLeuHisGluGlyGluGln 271
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1107 ATCAAAAAACAAAAGTTAATGTAAACCATTCATGATTAATGAAGGTTACAG 1156

272  TyrArgPheGlyClnThrGlnPheLeuGlyAsnLeuThrTyrThrGln 288
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1157 TATGACCTTCGTAGTCGACCAATTATAGGTAATTCGGAGGATGATGTCG 1206

288  aGluLeuGluAlaLeuLeuLys...PheLysAlaGluGluGlyPheSer 304
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

```

```

||||| 1207 CGAGCTTGAACCTTTACTTTTCAGCATTACATTTAAATGATACCTTTCGCC 1256
||||| 304 lnAlaMetLeuGluGlnThrThraSnAsnIleSerThrLysPheGlyAsp 320
||||| 1257 GTAGTGATATTGCAGATGTGCAAAATGCAATTAAGCAAACTTTGGAGAA 1306
||||| 321 AspGlyTyrTyrTyrAlaGlnIleArgProValThrArgIleAsnAspG1 337
||||| 1307 CGCGGTTACGGTACGCCAACGGTAATTCAGTACCTGATTTTGATGATGC 1356
||||| 337 uSerArgThrValAspValGluTyrTyrIleAspProValHisProValT 354
||||| 1357 AAATAAAACATATTACGATACACCTTGTGTGTGATGCTGGACGAGCTTAA 1406
||||| 354 yrValArgArgIleAsnPhethrGlyAsnPhelystThrGlnAspGluVal 370
||||| 1407 CTGTTCCCAACTTCGCTTTGAAGAAATACCGTTTCTGCTGATGACCT 1456
||||| 371 LeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLy 387
||||| 1457 TTAGCTCAGGAATGCCCAACAGAGGAACTTGGTATATTCACAATT 1506
||||| 387 sIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePheLysHisV 404
||||| 1507 AGTTGAGTTAGGAAAAATTCGCTTAGATCGTACAGGTTTCTTCGAA... 1552
||||| 404 alThrValAspThrArg.....ProValProAsnSerProAspGlnVal 418
||||| 1553 ..ACAGTCGAAAACCGAATTCATCTCAATGATGATGATGATGATGATG 1600
||||| 419 AspValAsnPhelValValGluGlnProSerGlySerThrIleAl 435
||||| 1601 GATGCTGATATAAAGTCAAGAACGTAACACGGGTAGTATCAACTTTGG 1650
||||| 435 aAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSerG 452
||||| 1651 TATTGGTTAGGGTACAGAGAGTGTATTAGTTATCAAGCAAGTGTAAAC 1700
||||| 452 lnAsnAsnPhemMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468
||||| 1701 AAGATAATTTCTGGGAACAGGGCGCGCATTAAGTATAGCTGGTACGAAA 1750
||||| 469 SerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheTh 485
||||| 1751 AATGATTATGTCAGAGTGCAATTTGGGTATACCGAGCCCTATTTTAC 1800
||||| 485 rValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLysT 502
||||| 1801 TAAAGATGGTCTA.....AGTCTTGGTGGAAATGTTTCTTTGAAACT 1844
||||| 502 yrAspAsnLys.....AsnIleSerAsnTyrValLeuAspSerTyr 515
||||| 1845 ACGATAACTCTAAAGTGATACATCCCTCTAATATATAAGCGTACGATAC 1894
||||| 516 GlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSe 532
||||| 1895 GGAAGTAATGTTACTTTAGGTTCCCTGTAAATGAAATAAATCACTCTATTA 1944
||||| 532 rPheGlyLeu.....AsnAlaAspAsnThrLysLeuHisG 544
||||| 1945 TGTAGGATTAGGTCATACCTATAATAAAATAGTAACCTTTGCTCTAGAA 1994
||||| 544 lyGlyArgPheMetGlyIleSerAsnValLysGlnLeuMetAlaAspGly 560
||||| 1995 ATACCGTAATTTATATTATCAATCAATGAAA..... 2026
||||| 561 GlyLysIleGlnValAspAsnAsnGlyIleProAspPheLysHisAspTy 577
||||| 2027 .....TTTAAAGGTAATGGCATTAACAACAATGACITTTGATTT 2064
||||| 577 rThrThrTyrAsnAlaIleLeuGlyTyrPAsnTyrSerSerLeuAspArgp 594
|||||

```

```

2065 TTCT.....TTTGGTTGGAACATAAACAGCCTTAATAGAG 2099
594 roValPheProThrGlnGlyMetSerHisSerValAsp.....LeuThr 608
||||| 2100 GCTATTTCCCAACTAAAGGGGTTAAAGCAAGTCTTGGTGGAGGATTACT 2149
||||| 609 ValGlyPheGlyAspLysThrHisGlnLysValValTyrGlnGlyAsnI1 625
||||| 2150 ATTCAGGTTCTGATATAACAATACTACAACTAAGTGCAGATGTACAGGG 2199
||||| 625 eTyrArgProPhe.....IleLysLysSerValL 635
||||| 2200 TTTTACCCATTAGACAGAGATCACCTCTGGGTTGTATCTGCAAAAGCAT 2249
||||| 635 euArgGlyTyrAlaLysLeuGlyTyrGlyAsn...AsnLeuProPheTyr 650
||||| 2250 CTGCAGGATATGCAAT...GGTTTGGAAACAACGGTTTACCGTTCAT 2296
||||| 651 GluAsnPhetYrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667
||||| 2297 CAAACTTATACAGCGGTGGCATCGGTTTCATTACGTGTTTTCGTTATGG 2346
||||| 667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
||||| 2347 TAGTATTGGACCTAACGCA.....ATTATGCCGAATATGTAATG 2387
||||| 684 lnThrThrLeuGly.....GluValValGlyGly 693
||||| 2388 GTAGTGTGCTAGCTGACTTTTAAAGAAAGATAAGTTCTGTGATGTTGGT 2437
||||| 694 AsnAlaLeuAlaThrPheGlySerGluLeuIleLeuProLeuProPheLy 710
||||| 2438 AATCAATCGCTACAGTACGAGCGAGGTTAATTGTGCCAACTCCATTGT 2487
||||| 710 sGlyAspTrpIle...AspGlnValArgProValIlePheIleGluGlyG 726
||||| 2488 GAGCGATAAGAGCCAAATAACGCTCCGCAACCTCTTATTGTTGATGCGG 2537
||||| 726 lyGlnValPheAspThrThr...GlyMetAspLysGlnThrIleAspLeu 741
||||| 2538 CAAGTGTTTGGATACTAAATGGAAATCAGATAAAATGGATAGAGAGC 2587
||||| 742 ThrGlnPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAl 758
||||| 2588 GATGATTAATAA..... 2599
||||| 758 aAsnArgProLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyV 775
||||| 2600 .AGATTGCTGATTATGGCAATCAAGCCGATTTCGCGCCCTCTACAGGTG 2648
||||| 775 alGlyAlaThrTyrTyrThrProIleGlyProLeuSerIleSerTyrAla 791
||||| 2649 TCGGATTCCAATGGCAATCTCTATTGGCCATTGGTATTCTTTATGCC 2698
||||| 792 LysProLeuAsnLysLysGlnAsnAspGlnThrAspThrValGlnPheG1 808
||||| 2699 AAACCAATTAATAAATATGAAATGATGATGTCGCAACAGTTCCTCAATTAG 2748
||||| 808 nIleGlySerValPhe 813
||||| 2749 TATTGGAGGTTCTTTC 2764
seq_name: /cnp2_6/ptodata/2/ina/6B_COMB.seq.us-08-942-046-3
seq_documentation_block:
; Sequence 3, Application US/08942046
; Patent No. 6264954
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles

```

```

alignment_scores:      Quality: 1052.50      Length: 822
                       Ratio: 1.956        Gaps: 20
Percent Similarity: 65.450      Percent Identity: 31.630

alignment_block:
US-09-701-711-2 x US-08-942-046-3      ..

Align seg 1/1 to: US-08-942-046-3 from: 1 to: 2984

23 SerThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIleThrIleTh 39
   :::::  ::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
413 ACACAGCAGTGTGTGGCGGCACCTTTTGTGCCAAAGATATCGGTGGA 462

39 rGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPheA 56
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
463 TGGTGTTCAGAGGTGACATTAGAACACAAATCCGAGCAAGTTTACCTGTT 512

56 rGlyLeuGlnValValSerGluAsnGlnLeuAlaAspGlyValLysAla 72
   || |||||  |||:  :::::  :::::  :::::  :::::  :::::  :::::  :::::
513 GTCCCGGTCAACGCTGTGACTGCACATGATGTGGCTTAATATTGTCGCTCT 562

73 LeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisGlnGlu 89
   ||||:  :::::  ||||:  ||||:  ||||:  ||||:  ||||:  ||||:  ||||:  ||||:
563 TTTATCTGAAGTGGTGCATTCGATGATGTGAAGCG...CATCAAGAAGG 609

89 yArGile...IleTyrGlnValThrGluArgProLeuLeAlaGluIleA 105
   - - - - -  - - - - -  - - - - -  - - - - -  - - - - -  - - - - -
610 CGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 659

```





```

; LENGTH: 2987 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-433-522A-55

```

```

alignment_scores:
  Quality: 1052.50      Length: 822
  Ratio: 1.956          Gaps: 20
  Percent Similarity: 65.450 Percent Identity: 31.630

```

alignment\_block:  
US-09-701-711-2 x US-08-433-522A-55

Align seg 1/1 to: US-08-433-522A-55 from: 1 to: 2987

```

23 SerThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIleThrIleTh 39
   ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
428 ACAAGACTGTGTTGGCGCACCTTTTGGCAAAAGATATTTCGTGTGA 477
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
39 rGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPheA 56
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
478 TGGCTGTCAAGTGACTTAGAACACAAATCGAGCAAGTTTACCTGTC 527
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
56 rGlyLeuGlnValValSerGluAsnGlnLeuAlaAspGlyValLysAla 72
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
528 GTGCGGTGAGCTGTGACTGACATGATGATGCGTAAATTTTCGCTCT 577
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
73 LeuTyAlaThrGlyAsnPheSerAspValGlnValTyHisGlnGluGl 89
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
578 TTATTCGTAAGTGGTCGATTCGATGATGTAAGCG...CATCAAGAAGG 624
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
89 yArgIle...IleTyGlnValThrGluArgProLeuIleAlaGluIleA 105
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
625 CGATGTGCTTGTGTGTAGCGTTGTGGCTAAATCGATCATTTTCAGAT 674
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
105 snPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGlyLeu 121
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
675 AAATCAAGAGTAACTCTGTATTCCTCCAGCACTTAACAAACTTAA 724
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
122 LysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrValG 138
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
725 GATGCTAAACGGGTTTAAAGTTCGCGATGTTTAAATTCGAGAAATTA 774
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
138 nMetIleGluThrGluLeuThrAsnGlnTyIleSerGlnGlyTyTyA 155
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
775 TGAATTTGCCAAAGTGTAAAGAGGACCTATGCCAAGTGTAGTGTCTA 824
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
155 snThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgValLys 171
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
825 ACCCAACAGTTGAACCTATTGTCAATACGCTACCAATAATCGCGTGA 874
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
172 LeuAspMetThrPheAlaGluGlyLysProAlaArgValValAspLeas 188
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
875 ATTTTAATCAAAATCAATGAAGATATAAGCAAAATTTGGCATCAAT 924
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
188 nIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspValLeu 205
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
925 TTTCAAGGGAGCAATCTCTTAGTACGACATACATTACAAGAACAAATG 974
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
205 latLeuAspAsnLysIleAsnProLeuSerLysAlaAspArgTyThr 221
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
975 AATTACAACTGATCTTGTGGTGAATTA...TGGGAAATAAATTTGAA 1021
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
222 GlnGluLysLeuValThrSerLeuGluAsnLeuArgAlaLysTyLeuAs 238
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1022 GTGCGCAATTCGAGAAGATTTGACGATCAATTCGTCATTATTATTA 1071
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
238 nAlaGlyPheValArgPheGluIleTyAspAlaLysLeuAsnIleAsnG 255
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1072 TAATGGCTATGCCAAACGACAAATTTACTAAACGGATGTTTCAGCTAA 1121
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

255 luAspLysAsnArgIlePheValGluIleSerLeuHisGluGlyGluGln 271
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1122 ATGAAAAACAAAAGTTAATGTAACTATGATGTAATGAAGTTTACAG 1171
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
272 TyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuThrTyThrGlnAl 288
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1172 TATGACCTTCGTAGTCACGATATTATAGTATCTGGGAGGTATGTCTGC 1221
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
288 aGluLeuAlaLeuLeuLys...PheLysAlaGluGlyPheSerG 304
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1222 CGAGCTTGAACCTTTACCTTCAGCATTAATTAATGATACTTTCCGCC 1271
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
304 InAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAsp 320
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1272 GTAGTGATATTGCAGATGTAGAAAATGCAATTAAGCAAAACTTCGAGA 1321
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
321 AspGlyTyThrTyThrAlaGlnIleArgProValThrArgIleAsnAspG 337
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1322 CGCGTTTACGGTAGCGCAACGGTAAATTCAGTACCTGATTGTGATGAT 1371
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
337 uSerArgThrValAspValGluTyThrTyThrIleAspProValHisPro 354
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1372 AAATAAACATTTAGCGATAACCCCTTTGTTGATGCTGGACGACGTTAA 1421
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
354 yValArgArgIleAsnPheThrGlyAsnPheLysThrGlnAspGluVal 370
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1422 CTGTTCCGCAACTTCGCTTGAAGAAATACCGTTTCTGCTGATGACACT 1471
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
371 LeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLy 387
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1472 TTACGTGAGGAATTCGCCCAACAGAGAACTTGTGTATAATTCACAATT 1521
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
387 stIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPheLysHisV 404
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1522 AGTTGAGTTAGGAAAAATTCGCTTAGCTACAGGTTTCTTCGAA... 1567
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
404 alThrValAspThrArg...ProValProAsnSerProAspGlnVal 418
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1568 ..ACAGTCGAAACCGAATTTGATCTCAATGGTAGTATGATGATGAAGTG 1615
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
419 AspValAsnPheValValGluGlnProSerGlySerSerThrIleAl 435
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1616 GATGCTGATATATAAGTCAAGAACGCTAACACGGGTAGTATCAACTTGG 1665
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
435 aAlaGlyTyThrSerGlnSerGlyValThrPheGlnPheAspValSerG 452
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1666 TATTGCTTACGCTACAGAGTGTATTAGTTATCAAGCAAGTGTAAAC 1715
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
452 InAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1716 AGATTAATTTCTTGGGAACAGGCGGCGGAGTAAAGTATAGCTGTACGAA 1765
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
469 SerGluThrArgGluValTyThrSerLeuGlyMetThrAsnProTyPheth 485
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1766 AATGATTATGTTACGAGTGTCAATTTGGGTTATACCGAGCCCTATTATT 1815
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
485 rValAsnGlyValSerGlnSerLeuSerGlyTyThrArgLysThrLys 502
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1816 TAAAGATGGTGA...AGTCTTGTGGAAATGTTTTCTTTCAAAACT 1859
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
502 yAspAsnLys...AsnIleSerAsnTyThrValLeuAspSerTyThr 515
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1860 ACATTAATCTTAAAGTATACATCTCTTAATATAAGCGTACGACTTAC 1909
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
516 GlyGlySerLeuSerTyThrGlyTyThrProIleAspGluAsnGlnArgIleSe 532
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1910 GGAAGTAATGTTACTTTTAGGTTTCCCTGTAAATGAAATAACTCCTATTA 1959
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
532 rPheGlyLeu...AsnAlaAspAsnThrLysLeuHisG 544
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1960 TGTAGGATTAGTCAATACCTATATAAATAATAGTAACCTTTGCTCTAGAAT 2009
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
544 lyGlyArgPheMetGlyIleSerAsnValLysGlnLeuMetAlaAspGly 560
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```



[illegible]

```

:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2362 TAGATTGACCTAAGCGCA.....ATTATGCGGAATATGGTAATG 2402
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
684 lNThrThrLeuGly.....GluValValGlyGly 693
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2403 GTAGTGGTACTGCTACTTTTAAAGACAGATAAGTCTGATGTGATGGTGT 2452
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
694 AsnAlaLeuAlaThrPheGlySerGluLeuLeuLeuProLeuProPheLeu 710
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2453 AATGCAATCGCTACAGTACGACAGTAAATGTCGCCAACCTCCATTGT 2502
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
710 sGlyAspTrpIle...AspGlnValArgProValIlePheIleGluGlyG 726
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2503 GAGCGGATAGAGCCCAAAATACGCTCCGAACCTCTTATTGTGTGTCGG 2552
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
726 lYcGlnValPheAspThrThr...GlyMetAspLysGlnThrIleAspLeu 741
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2553 CAAGTGTGTGGTAACTAAATGAAATCAGATAAATAAGATGATAGAGCC 2602
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
742 ThrGlnPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAl 758
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2603 GATGTATTAAAA..... 2614
758 asnArgProLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyV 775
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2615 AGATTGCTGATTATGCAAAATCAAGCGGTATTGCGCCTCTACAGGTG 2663
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
775 alGlyAlaThrTrpTyrThrProIleGlyProLeuSerIleSerTyrAla 791
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2664 TCGGATTCATGGAACATCTCTATTGGCCCATTTGGTATCTCTTATGCC 2713
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
792 LysProLeuAsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGl 808
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2714 AAACCAATTTAAAAATATCAAAATGATGATGTCGAACAGTTCCAATTAG 2763
808 nIleGlySerValPhe 813
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2764 TATTGGAGGTTCTTTC 2779
```

seq\_name: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.us-08-942-046-55

```

seq_documentation_block:
; Sequence 55, Application US/08942046
; Patent No. 6264954
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/942,046
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-732 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 55:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2987 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-942-046-55

alignment_scores:
    Quality: 1052.50      Length: 822
    Ratio: 1.956          Gaps: 20
    Percent Similarity: 65.450      Percent Identity: 31.630

alignment_block:
US-09-701-711-2 x US-08-942-046-55 ..
Align seg 1/1 to: US-08-942-046-55 from: 1 to: 2987

23 SerThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIleThrIleTh 39
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
428 ACAACGACGTGTGTGGCCGACCTTTTGTGGCAAAAGATATTCGTGTGGA 477
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
39 rGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPheA 56
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
478 TGGTGTCAAGGTGACTTAGAACACAACAAATCCGAGCAAGTTTACCTGTT 527
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
56 rGlyGlyGlnValValSerGluAsnGlnLeuAlaAspGlyValLysAla 72
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
528 GTGCGCGTGCAGCGTGTGACTGACATGATGATGGCTAATATTTGCCGCT 577
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
73 LeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisGlnGluG 89
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
578 TTATTCGTAAAGTGTGCTGATGATGATGTAAGCG...CATCAAGAGG 624
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89 yArgIle...IleTyrGlnValThrGluArgProLeuIleAlaGluIleA 105
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
625 CGATGTGCTTGTGTAGCGTTGTGGCTAAATCGATCATTTTCAGATGTTA 674
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
105 snPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluGlyLeu 121
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
675 AAATCAAGGTAACTCTGTTATTCCCACTGAAGCACTTAACAAAACTTA 724
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
122 LysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrValGl 138
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
725 GATGCTAACGGGTTAAAGTTGGCGATGTTTAATTCGAGAAAAATATAA 774
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
138 nMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyrTyrA 155
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
775 TGAATTTGCCAAAAGGTAAAGAGACGACATGCAAGTGTAGTGCCTATA 824
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
155 snThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgValLys 171
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
825 ACGCAACAGTTGAACCTATTGTCAATACGCTACCAATAATCGCGCTGAA 874
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
172 LeuAspMetThrPheAlaGluGlyLysProAlaArgValValAspIleAs 188
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
875 ATTTTAATTCAAATCAATGAAGATGATAAAGCAAAATTCGCATCAATAA 924
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
188 nIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspValLeuA 205
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
925 TTTCAAGGGGAACGAATCTGTTAGTAGCAGTACATTACAAGAACAATGG 974
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
205 laIleLysAspAsnLysIleAsnProLeuSerLysAlaAspArgTyrThr 221
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
975 AATTACAACCTGATTCTTGGTGGAAATTA...TGGGAAATAAATTTGAA 1021
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

[illegible]

```
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2714 AAACCAATTAAAAATATGAAATGATGTCGACAGTTCCAATTAG 2763
808 nileGlySerValpHe 813
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2764 TATTGGAGTCTTTC 2779
```

seq\_name: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:US-08-433-522A-9

```
seq_documentation_block:
; Sequence 9, Application US/08433522A
; Patent No. 6013514
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433,522A
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-434 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2989 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 390..2768
; US-08-433-522A-9
```

```
alignment_scores:
  Quality: 1047.50      Length: 818
  Ratio: 1.947          Gaps: 21
  Percent Similarity: 65.770  Percent Identity: 31.663
```

```
alignment_block:
US-09-701-711-2 x US-08-433-522A-9
..
Align seg 1/1 to: US-08-433-522A-9 from: 1 to: 2989
23 SerThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIleThrIleTh 39
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
429 ACACGACACTGTTTCCCGCACCTTTTGGCCAAAGATATTCCTGTGGA 478
39 rGlyLeuGlnArgValThrIleGlySerLeuGlnSerValLeuProPheA 56
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
479 TGGTGTTCAGGTGACCTTAGAACCAACAAATCCGAGCAAGTTTACCTCTTC 528
56 rGlyLeuGlnValValSerGluAsnGlnLeuAlaAspGlyValLysAla 72
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
529 GTGCTGTCAGCGTGTGACTGACACATGATGGCTAATATTTGTCGCTCT 578
73 LeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisGlnGluG 89
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
579 TTATTCTGAAGTGGTCGATCGATGATGTAAGCG...CATCAAGAAGG 625
89 yArgIle...IleTyrGlnValThrGluArgProLeuIleAlaGluIleA 105
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
626 CGATGTCGCTTGTGTTAGCGTTGCTGCTAAATCGATCATTTAGATGTTA 675
105 snPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluGlyLeu 121
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
676 AATCAAGGTAACCTCTGTTATTTCCCACTGAGCACCTTAACAAACTTA 725
122 LysAsnAlaGlyLeuAlaValGlyGlnProLysGlnAlaThrValG 138
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
726 GATGCTAACGGGTTTAAAGTTGGCGATGTTTAAATTCGAGAAAAATAAA 775
138 nMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyrTyrA 155
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
776 TGAATTTGCCAAAAGTGTAAAGAGACACTATGCAAGTGTAGGTCGCTATA 825
155 snThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgValLys 171
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
826 ACGCAACCGTTGAACCTATTGTCATAGCTGCCAAATAATATCGTGCCTGAA 875
172 LeuAspMetThrPheAlaGluGlyLysProAlaArgValValAspIleAs 188
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
876 ATTTTAAATCAATCAATGAAGATGATAAGCAAAATTTGCATCATTAAC 925
188 nIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspValLeuA 205
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
926 TTTCAAGGGGAGCAATCTGTTAGTAGCAGTACATTACAAGAACAAATGG 975
205 laIleLysAspAsnLysIleAsnProLysSerLysAlaAspArgTyrThr 221
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
976 AATTACAACCTGATCTTGGTGGAATTA...TGGGGAAATAAAATTTGAA 1022
222 GlnGluLysLeuValThrSerLeuGluAsnLeuArgAlaLysTyrLeuAs 238
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1023 GGTGCGCAATTCGAGAAAGATCTCGAGCAATTCGTGATTATTATTTAAA 1072
238 nAlaGlyPheValArgPheGluIleLysAspAlaLysLeuAsnIleAsnG 255
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1073 TAATGGCTATGCCAAAGCACAAATCACTAAACGGATGTTTCAGCTAAATG 1122
255 luAspLysAsnArgIlePheValGluIleSerLeuHisGluGlyGluGln 271
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1123 ATGAAAAACAAAAGTTAATGTACCATTTAGTAAATGAAGGTTTACAG 1172
272 TyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAl 288
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1173 TATGACCTTCGTAGTCACCAATTATAGTAATCTGGAGGATGCTCTGC 1222
288 aGluLeuGluAlaLeuLys...PheLysAlaGluGluGlyPheSerG 304
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1223 CGAGCTTGAACCTTTACTTTCAGCATTTACATTTAAATGATACTTTCCGCC 1272
304 lnAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAsp 320
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1273 GTAGTGATTTGCAGATGTAGAAAATCCAATTAAGCAAACTTGGGGAA 1322
321 AspGlyTyrTyrThrAlaGlnIleArgProValThrArgIleAsnAspGl 337
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1323 CGAGGTTCACGTAACACACAGATAAATCTGTACCTCATTTTGACGATGC 1372
337 uSerArgThrValAspValGluTyrTyrIleAspProValHisProValT 354
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1373 AAATAAACATTAGCGATAACCTTTGTTGTTGATGCTGGACGACGTTTAA 1422
```

Thu Sep 19 10:04:26 2002

```

354 yrvAlaArgIleAsnPhetHrGlyAsnPhelYsThrGlnAspGluVal 370
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1423 CTGTTGGCAACTTCGCTTGAAGAAATACCGTTTCCTGCTAGTAGTACT 1472
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
371 LeuArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLy 387
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1473 TTACGTCAGGAAATCGCACAAAGAAAGCACTTGGTATATAATTCACAATT 1522
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
387 stleGlnLeuSerArgAlaArgLeuMetArgHrGlyPhePheLysHisv 404
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1523 AGTTAGTTAGGAAATTCGTTAGATCGTACAGGTTTCTTCGAA... 1568
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
404 alThrValAspThrArg.....ProValProAsnSerProAspGlnVal 418
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1569 ..ACAGTTGAAACCGAATGTATCTTCATCAATGGTAGCAATGATGAAGTG 1616
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
419 AspValAsnPhelValGluGluGlnProSerGlySerSerThrIleAl 435
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1617 GATGTCGTATATAAGTCAAGAACGTAACACGGGTAGTATCACTTTGG 1666
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
435 aLaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSerG 452
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1667 TATGTTAGGTCACAGAGAGTGTATCAGTTATCAAAACAAGTATTAAAC 1716
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
452 lnAsnAsnPhetMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1717 AAGATATTTCTTGGGAACAGGGCGCGCAGTAGTAGTACGTACGAA 1766
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
469 SerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheTh 485
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1767 AATGATTTATGTCAGGTGTCAATTTGGTTATACCGAACCCCTATTATTAC 1816
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
485 rValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLysT 502
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1817 TAAAGATGGTGTA.....AGTCTTGTGTGGAATATTTCTTTGAAAACT 1860
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
502 yAspAsnLys.....AsnIleSerAsnTyrValLeuAspSerTyr 515
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1861 ACGATACTCTAAAGATGATACCTCTACATATAGCGTACGACTAT 1910
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
516 GlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSe 532
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1911 GGAAGTAATGTTACTTTAGGTTTCCCTCTAAATGAAATTAACCTCTATTA 1960
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
532 rPheGlyLeu.....AsnAlaAspAsnThrLysLeuHisG 544
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1961 TGTAGGATAGCGCCATACCTATAATAAATTAGTAACCTTTGCTCTAGAAAT 2010
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
544 LyGlyArgPheMetGlyIleSerAsnValLysGlnLeuMetAlaAspGly 560
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2011 ATAACCGTAATTTATATATATCAATCAATGAAA..... 2042
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
561 GlyLysIleGlnValAspAsnAsnGlyIleProAspPheLysHisAspTy 577
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2043 .....TTTAAAGGTAATGGCATTAACAAAGTACTTTGATTT 2080
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
577 rThrThrTyrAsnAlaIleLeuGlyTyrAsnTyrSerSerLeuAspArgp 594
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2081 TTCT.....TTTGGTTGGAACATAACACGCTTATATAGAG 2115
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
594 roValPheProThrGlnGlyMetSerHisSerValAsp.....LeuThr 608
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2116 GCTATTTCCCAACTAAGGGTTAAAGCAAGTCTTGTGGACGAGGTACT 2165
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
609 ValGlyPheGlyAspLysThrHisGlnLysValVal.....TyrGlnGl 623
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2166 ATTCCAGGTTCTGATAACAAATACACAACTAAGTGCAGATGTACAGG 2215
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
623 y.....AsnIleTyrArgProPheIleLysLysSerValL 635
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2216 TTTCACCATTTAGACAGAGATCACCGCTGGTGTGTATCTGCAAAAGCAT 2265
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

seq\_name: /cgn2\_6/ptodata/2/lna/6A\_COMB.seq:US-09-135-166-9

```

seq documentation block:
; Sequence 9, Application US/09135166
; Patent No. 6083743
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```



```

; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/135.166
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2989 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 390..2768
; US-09-135-166-9

```

## alignment\_scores:

```

    Quality: 1047.50      Length: 818
    Ratio: 1.947          Gaps: 21
    Percent Similarity: 65.770      Percent Identity: 31.663

```

## alignment\_block:

```
US-09-701-711-2 x US-09-135-166-9 ..
```

```
Align seg 1/1 to: US-09-135-166-9 from: 1 to: 2989
```

```

23 SerThrHisAlaGlnAlaAlaPheMetAlaAsnAspIleThrIleTh 39
429 ACAACGACTGTGTTCCCGCACCCTTTGTGCCAAAGATATTCGTGGA 478
39 rGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPhe 56
479 TGGTGTTCAGGTGACTTAGAACACAAATCCGAGCAAGTTTACCTGTC 528
56 rGlyGlyGlnValValSerGluAsnGlnLeuAlaAspGlyValLysAla 72
529 GTGCTGGTCAGCGTGTGACTGACAAATGATGGCTAATATGTCCGCTCT 578
73 LeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisGlnGlu 89
579 TTATTCGTAAAGTGGTCGATCGATGATGTAAGCG...CATCAAGAAG 625
89 yArgIle...IleTyrGlnValThrGluArgProLeuIleAlaGluIle 105
626 CGATGCGCTGTTGTTAGCGTTTGGCTTAATTCGATCAATTCAGATGTA 675
105 snPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluGly 121
676 AAATCAAGGTAACTCTGTTATTCACCTGAAGCACTTAACAAAACTTA 725
122 LysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrVal 138
726 GATGCTAACGGGTTTAAAGTTGGCGATGTTTAAATTCGAGAAAAATTA 775
138 nMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyrA 155
776 TGAATTTGCCAAAGGTAAAGAGACACTATGCAAGCTGTAGTGCCTATA 825
155 snThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgVal 171

```

```

826 ACGCAACCGTTGAACCTATTGTCAATACGCTGCCAAATAATCGTGCTGA 875
172 LeuAspMetThrPheAlaGluGlyLysProAlaArgValValAspIleAs 188
876 ATTTTAATTCAAAATCAATGAAGATGATAAGCAAAATTTGCCATCATTA 925
188 nIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspValLeu 205
926 TTTCAGGGGGAACGAATCTCTTAGTAGCAGTACATTACAGAACAATGG 975
205 laIleLysAspAsnLysIleAsnProLeuSerLysAlaAspArgTyrThr 221
976 AATTACAACCTGATCTCTTGGTGGAAATTA...TCGGGAAATAAATTTGA 1022
222 GlnGluLysLeuValThrSerLeuGluAsnLeuArgAlaLysTyrLeuAs 238
1023 GTGCGCAATTCGAGAAAGATCTGCAGGCAATTCGTGATTATTATTAAA 1072
238 nAlaGlyPheValArgPheGluIleLysAspAlaLysLeuAsnIleAsn 255
1073 TAATGCTATGCCAAAGCACAAATCACTAAACGGATGTTTCAGCTAAATG 1122
255 luAspLysAsnArgIlePheValGluIleSerLeuHisGluGlyGln 271
1123 ATGAAAAACAAAAGTTAATGTAACTTACCATTTGATGTAATGAAGTTTACAG 1172
272 TyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAl 288
1173 TATGACCTTCGTAGTCACGCAATTATAGGTAATCTGGGAGGTATGCTGC 1222
288 aGluLeuGluAlaLeuLeuLys...PheLysAlaGluGluGlyPheSerG 304
1223 CGAGCTTGAACCTTTACTTTCAGCATTTACATTTAAATGATACTTTCGCC 1272
304 lnAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAsp 320
1273 GTAGTGATATTGCAGATGTAGAAAAATGCAATTAAGCAAACTTGGGGAA 1322
321 AspGlyTyrTyrThrAlaGlnIleArgProValThrArgIleAsnAspG 337
1323 CGAGGTTCGGTAACACACAGTAATAATCTGTACCTGATTTTCACGATGC 1372
337 uSerArgThrValAspValGluTyrTyrIleAspProValIHisProVal 354
1373 AAATAAACAATTAGCATACCTTTGTTGTGTGATGCTGGACGACGTTAA 1422
354 yrValArgArgIleAsnPheThrGlyAsnPhelyThrGlnAspGluVal 370
1423 CTGTTCCCAACTTCGCTTTGAAGGAAATACCGTTTCTGCTGATAGTACT 1472
371 LeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLy 387
1473 TTACGTCAGGAATGCGACACACAAAGGAAGAACTTGGTATTAATTCACA 1522
387 sIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePheLysHis 404
1523 AGTTGAGTTAGGAAAAATTCGCTTAGATCGTACAGGTTTCTTCGAA... 1568
404 aThrValAspThrArg.....ProValProAsnSerProAspGlnVal 418
1569 ..ACATGAAAAACCGAATTTGATCCTTATCAATGGTAGCAATGATGAAGT 1616
419 AspValAsnPheValValGluGluGlnProSerGlySerSerThrIleAl 435
1617 GATGTCGTATATAAGTCAAGAACGTAACCGGCTAGTAGTATCAACTTGG 1666
435 aAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSerG 452
1667 TATTGCTTACGCTACAGAGAGTGGTATCAGTTTATCAACAAGTATTAAAC 1716
452 lnAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468
1717 AAGATAATTTCTTTGGGAACACGGGGCGCGCTAAGTATAGCTGTGACGAA 1766

```

```

469 SerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheTh 485
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1767 AATGATTATGGTACGAGTGCTAAATTTGGGTATATACCGAACCTATTATTC 1816
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
485 rValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLysT 502
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1817 TAAAGATGGTGTA.....AGTCTTGGTGGAAATATTTCTTTGAAAACT 1860
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
502 YrAspAsnLys.....AsnIleSerAsnTyrValLeuAspSerTyr 515
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1861 ACGATAACTCTAAAAGTGATACATCTCTAATACTATAAGCGTACGACTTAT 1910
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
516 GlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSe 532
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1911 GGAAGTAAATGTTACTTTAGTGGTTCCTCTAAATGAAATAACTCTCTATTA 1960
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
532 rPheGlyLeu.....AsnAlaAspAsnThrLysLeuHisG 544
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1961 TGTAGGATTAGCCCATACCTATATAATAAATTAGTAACCTTTGCTCTAGAAT 2010
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
544 IyGlyArgPheMetGlyIleSerAsnValLysGlnLeuMetAlaAspGly 560
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2011 ATAACCGTAATTATATATATCAATCAATCAAA..... 2042
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
561 GlyLysIleGlnValAspAsnGlyIleProAspPheLysHisAspTy 577
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2043 .....TTTAAAGGTAATGGCACTTAAACAAATGACTTTGATTT 2080
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
577 rThrThrTyrAsnAlaIleLeuGlyTyrAsnTyrSerSerLeuAspArgP 594
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2081 TTCT.....TTTGGTTGGAACATAACAGCCTTAATAGAG 2115
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
594 rValPheProThrGlnGlyMetSerHisSerValAsp.....LeuThr 608
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2116 GCTATTTCCTCAACTAAAGGGTAAAGCAAGTCTTGGTGGAGAGTTACT 2165
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
609 ValGlyPheGlyAspLysThrHisGlnLysValVal.....TyrGlnG 623
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2166 ATCCAGGTTCTGTAAACAATACTACAACACTAAGTCGAGATGACAGGG 2215
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
623 Y.....AsnIleTyrArgProPheLeuLysSerValL 635
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2216 TTCTACCATTTAGACAGAGATCACCGCTGGTGTATCTATCGCAAAAGCAT 2265
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
635 euArgGlyTyrAlaLysLeuGlyTyrGlyAsn....AsnLeuProPheTyr 650
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2266 CTGCAGGATATGCAAT...GTTTGGAAACAAGCGTTTACCGTTCTAT 2312
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
651 GluAsnPheTyrAlaGlyTyrGlySerValArgGlyTyrAspGlnSe 667
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2313 CAAACTATACAGCGGTGGCATGTTGTTCTATACGCGGTTTTCATTATGG 2362
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 584
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2363 TAGATTGGGCTCAATGCA.....ATTTATGCCGAACATGTTGTAATG 2403
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
684 InThr.....ThrLeuGlyGluValValGlyAsnAlaLeuAla 697
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2404 GTCATTTTAAATAGATAAGTTCGTGATGATGTTGGTGAATGCAATCACA 2453
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
698 ThrPheGlySerGluLeuLeuProLeuProPheLysGlyAspTrpI 714
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2454 ACTGCAGTGCAGAACTTATTGTACCAACTCCATTTGTGAGTGATAAAG 2503
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
714 e....AspGlnValArgProValIlePheIleGlyGlyGlnValPhe 730
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2504 CCAAAATACAGTCCCAACCTCCCTATTGTGTGATGCGCAAGTGTGGA 2553
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
730 spThrThr...GlyMetAspLysGlnThrIleAspLeuThrGlnPheLys 745
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2554 ATACTAAATGGAATCAGATAAATAATGGATTAGAGAGCAAGGCTTTGAAA 2603
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

```

```

746 AspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLe 762
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2604 GAC.....TTACCTGA 2614
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
762 uLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrT 779
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2615 TTATGGCAAAATCAAGCGTATTCGCCCTCTACAGGTGTCGGATTCCAAT 2664
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
779 rPtyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeuAsn 795
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2665 GCGAATCTCTATTGGACCATTTGTTTCTTATGCTAAACCAATATAA 2714
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
796 LysLysGlnAsnAspGlnThrAspThrValGlnPheGlnIleGlySerVa 812
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2715 AAATATGAAATGATGATGTGCAACAGTTCCAATTTAGTATTGGGGCTC 2764
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
812 lPhe 813
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2765 TTTC 2768
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-942-046-9

seq_documentation_block:
; Sequence 9, Application US/08942046
; Patent No. 6264954
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuen Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/942,046
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/433,522
FILING DATE: 12-SEP-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-732 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2989 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 390...2768
US-08-942-046-9

```

```
alignment_scores:
  Quality: 1047.50      Length: 818
  Ratio: 1.947          Gaps: 21
  Percent Similarity: 65.770  Percent Identity: 31.663

alignment_block:
US-09-701-711-2 x US-08-942-046-9

Align seg 1/1 to: US-08-942-046-9 from: 1 to: 2989

23 SerThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIleThrIleTh 39
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
429 ACAACGACTGTGTTGGCCGACCTTTGGCCCAAGATATTCGTGTGA 478
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

39 rGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPheA 56
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
479 TGGTGTCAAGTGACTTAAAGCAACAATCCGAGCAAGTTTACCTGTTTC 528
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

56 rGlyLeuGlnValValSerGluAsnGlnLeuAlaAspGlyValLysAla 72
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
529 GNGCTGGTCAGCGTGTGACTGACATGATGCTGCTAAATTTGTCGCTCT 578
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

73 LeuTyAlaThrGlyAsnPheSerAspValGlnValTyHisGlnGluG 89
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
579 TTATTCGTAAAGTGTGATTCGATGATGAAAGCG...CATCAAGAGG 625

89 yArgile...IleTyGlnValThrGluArgProLeuIleAlaGluIleA 105
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
626 CGATGCTGTGTTGGTGTAGCGTTTGGCTAAATTCGATCATTTTCAGATG 675

105 snPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluGlyLeu 121
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
676 ARAATCAAGGTAACTCTGTATTCCTGAGCACTTAACAAACATTA 725

122 LysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrValG 138
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
726 GATGCTAACGGGTTTAAAGTTGGCGATGTTTAAATTCGAGAAATTA 775

138 nMetIleGluThrGluThrAsnGlnTyIleSerGlnGlyTyTyra 155
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
776 TGAATTTGCCAAAGGTAAAGACGACATCAAGGTAGTGTGCTGCTATA 825

155 snThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgValLys 171
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
826 ACACCAACGGTTCAACCTATTGTCAATACGTCGCCAAATAATCGTGTGA 875

172 LeuAspMetThrPheAlaGluGlyLysProAlaArgValValAspIleAs 188
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
876 ATTTTAATTCAAATCAATGAAGATGATAAAGCAAAATTTGGCATCATTA 925

188 nIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspValLeu 205
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
926 TTTCAAGGGGAACGAACTGTTAGTAGCAGTACATTACAGAACAAATGG 975

205 laIleLysAspAsnLysIleAsnProLeuSerLysAlaAspArgTyThr 221
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
976 AATTACAACCTGATCTTGGTGGAAATTA...TGGGAAATAAATTTGAA 1022

222 GlnGluLysLeuValThrSerLeuGluAsnLeuArgAlaLysTyLeuAs 238
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1023 GGTGGCGCAATTCGAAAGATCTGAGGCAATTCGTGATTATTATTAA 1072

238 nAlaGlyPheValArgPheGluIleLysAspAlaLysLeuAsnIleAsnG 255
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1073 TAATGGCTATGCCAAGCAAAATCACTAAACGGATGTTTCAGCTAAATG 1122

255 luAspLysAsnArgIlePheValGluIleSerLeuHisGluGlyGln 271
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1123 ATGAAAAACAAAGCTTAATGTAACTGATGATGATAAATGAAGTTTACAG 1172

272 TyrArgPheGlyGlnValThrGlnPheLeuGlyAsnLeuThrTyThrGln 288
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
2043 .....TTTAAAGGTAAATGGCAATTAACAAACAAATGACTTTGATTT 2080
```

```
1173 TATGACCTTCGTAGTCCACGCATTATAGTAATCTGGGAGGTATGTCTGC 1222
288 aGluLeuGluAlaLeuLeuLys...PheLysAlaGluGluGlyPheSerG 304
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1223 CGAGCTTGAACCTTTACTTTTCAGCATTTACATTAATGATGACTTTCCGCC 1272
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

304 InAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAsp 320
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1273 GTAGTGATATTGCAGATGTAGAAATGCAATTAAGCAAAACTTTGGGAA 1322

321 AspGlyTyTyThrAlaGlnIleArgProValThrArgIleAsnAspG 337
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1323 CGAGGTACGGTAAACACACAGATAAATCTGTACCTGATTTTGACGATGC 1372
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

337 uSerArgThrValAspValGluTyTyThrIleAspProValHisProVal 354
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1373 AAATAAAACATTAGCGATAAATCTGTTGTTGATGCTGGACGACGTTTAA 1422

354 yfValArgArgIleAsnPheThrGlyAsnPheLysThrGlnAspGluVal 370
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1423 CTGTTCCGCAACTTCGCTTTGAAGGAATACCGTTTCTGCTGATAGTACT 1472

371 LeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGln 387
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1473 TTACGTCAGGAAATCGACAAAGAACTTGGTATAAATTCACAAT 1522

387 sIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePheLysHis 404
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1523 AGTTGAGTTAGGAAATTCGCTTAGATCGTACAGGTTTCTTCGAA... 1568

404 alThrValAspThrArg.....ProValProAsnSerProAspGlnVal 418
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1569 ..ACAGTTGAAACCGAATTTGATCTATCAATGGTAGCAATGATGAAGTG 1616

419 AspValAsnPheValValGluGlnProSerGlySerSerThrIleAl 435
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1617 GATGTCGTATAAAAGTCAAGAAAGCTAACACGGGTAGTATCAACTTTGG 1666

435 aAlaGlyTySerGlnSerGlyGlyValThrPheGlnPheAspValSerG 452
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1667 TATTTGTTACGGTACAGAGGTGATCAGTTATCAACAAAGTATTAAAC 1716

452 InAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1717 AAGATAAATTTCTTGGCAACAGGGCGCAGTAAGTATAGCTGGTACGAAA 1766

469 SerGluThrArgGluValTySerLeuGlyMetThrAsnProTyThrPhe 485
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1767 AATGATTATGGTACGAGTGTCAATTTGGGTTATACCGAACCCCTATTAC 1816

485 rValAsnGlyValSerGlnSerLeuSerGlyTyTyThrArgLysThrLys 502
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1817 TAAAGATGGTGA.....AGTCTTGGTGAATAATTTTCTTTGAAAAC 1860

502 yfAspAsnLys.....AsnIleSerAsnTyfValLeuAspSerTy 515
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1861 ACGATAACTCTAAAAGTGTATACATCTCTTAAGTATAAGCGTACGACTTAT 1910

516 GlyGlySerLeuSerTyGlyTyProIleAspGluAsnGlnArgIleSe 532
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1911 GGAAGTAATGTACTTTAGGTTTCCCTGTAAATGAAATAAATAACTCTATTA 1960

532 rPheGlyLeu.....AsnAlaAspAsnThrLysLeuHisG 544
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1961 TGTAGATTAGGCCCACTACCTATAATAAATTAATTAACCTTTGCTCTAGAAT 2010

544 lyGlyArgPheMetGlyIleSerAsnValLysGlnLeuMetAlaAspGly 560
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
2011 ATAACCGTAATTTATATATCAATCAATCAAA..... 2042

561 GlyLysIleGlnValAspAsnAsnGlyIleProAspPheLysHisAspTy 577
   |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
2043 .....TTTAAAGGTAAATGGCAATTAACAAACAAATGACTTTGATTT 2080
```

```
577 rThrThrTyrAsnAlaIleLeuGlyTyrAsnTyrSerSerLeuAspArgp 594
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2081 TTCT .....TTTGGTTGGAACATAACAGCCTTAATAGAG 2115
594 roValPheProThrGlnGlyMetSerHisSerValasp.....LeuThr 608
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2116 GCTATTTCCTCAACTAAAGGGGTTAAAGCAAGTCTTGGTGGACGAGTTACT 2165
609 valGlyPheGlyAspLysThrHisGlnLysVal.....TyrGlnG1 623
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2166 ATTCAGGTTCTGTATACAAATACTACAACTAAGTCACAGTGTACAGGG 2215
623 Y.....AsnIleTyrArgProPheIleLysLysSerValL 635
2216 TTTCTACCATAGACAGATCACCGCTGGTGTATCTGCAAAAGCAT 2265
635 euArgGlyTyrAlaLysLeuGlyTyrGlyAsn...AsnLeuProPheTyr 650
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2266 CTGCAGGATATGCAAT...GGTTTGGAAACAAGCGTTTACCGTCTCTAT 2312
651 GluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2313 CAAACTTATACAGCGGTGGCATTTGGTTCATTACGCGGTTTGGCTTATGG 2362
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2363 TAGTATGGGCTTAATGCA.....ATTATGCCGACATGGTAAAG 2403
684 lnThr.....ThrLeuGlyGluValValGlyGlyAsnAlaLeuAla 697
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2404 GTACTTTTAATAAGATAAGTTCTCATGTGATGGTGGTAAATGCAATCACA 2453
698 ThrPheGlySerGluLeuLeuLeuProLeuProPheLysGlyAspTrpI1 714
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2454 ACTCGAGTGCAGAACCTATTGTACCAACCTCCATTGTGAGTGATAAAAG 2503
714 e...AspGlnValArgProValIlePheIleGluGlyGlyGlnValPheA 730
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2504 CCAAAATACAGTCCGACCTCCCTATTGTGTGATGCGCAAGTGTTTGGA 2553
730 spThrThr...GlyMetAspLysGlnThrIleAspLeuThrGlnPheLys 745
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2554 ATACTAAATGGAATCAGATAAAATGGATTAGAGAGCAAGGTCTTGAAA 2603
746 AspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLe 762
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2604 GAC.....TTACCTGA 2614
762 uLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrT 779
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2615 TTATGGCAAAATCAAGCGGTATTCGCGCTCTACAGGTGTCGGATTCCAAT 2664
779 rPtyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeuAsn 795
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2665 GGCAATCTCTATTGGACCATTTGTTTCTTATGCTAAACCAATTTAA 2714
796 LysLysGlnAsnAspGlnThrAspThrValGlnPheGlnIleGlySerVa 812
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2715 AAATATGAAATGATGATGTCGACAGTTCCAATTTAGTATGGGGGCTC 2764
812 lPhe 813
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2765 TTTC 2768

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-433-522A-7
seq_documentation_block:
; Sequence 7, Application US/08433522A
; Patent No. 6013514
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
```

```
APPLICANT: YANG, Yan Ping
APPLICANT: LOOSMORE, Sheena
APPLICANT: SIA, Dwo Yuan Charles
APPLICANT: KLEIN, Michel
TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6TH Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/433,522A
FILING DATE: 12-SEP-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-434 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2974 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 386..2761
US-08-433-522A-7

alignment_scores:
Quality: 1042.00 Length: 819
Ratio: 1.940 Gaps: 20
Percent Similarity: 65.568 Percent Identity: 31.380

alignment_block:
US-09-701-711-2 x US-08-433-522A-7 ..
Align seg 1/1 to: US-08-433-522A-7 from: 1 to: 2974

23 SerThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIleThrIleTh 39
::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
425 ACAACGACTGTGTGGCCGACCTTTGTGGCAAAAGATATTCTGTGGA 474
39 rGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPheA 56
::: :::: :::: :::: :::: :::: :::: :::: ::::
475 TGTGTTCAGGTGACCTAGAACAAACAAATCCGAGCAAGTTTACCTGTC 524
56 rGlyGlyGlnValValSerGluAsnGlnLeuAlaAspGlyValLysAla 72
::: :::: :::: :::: :::: :::: :::: :::: ::::
525 GTCCCGTCAGCGGTGACTGACAAATGATGCTAATATTGTCGCTCT 574
73 LeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisGlnGluG1 89
::: :::: :::: :::: :::: :::: :::: :::: ::::
575 TTATTCGTAAGTGTGTCGATTCGATGATGTGAAGCG...CATCAAGAGG 621
89 YArgIle...IleTyrGlnValThrGluArgProLeuIleAlaGluLeA 105
::: :::: :::: :::: :::: :::: :::: :::: ::::
622 CGATGTGCTGTTGTTGTCGCTGTTGGCTAAATCGATCATTTACATGTTA 671
105 snPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluGlyLeu 121
```



```
697 AlaThrPheGlySerGluLeuLeuLeuProLeuProPheLysGlyAspThr 713
||||| : : : : : ||||| : : : : : ||||| : : : : :
2444 GCTACGCTAGCGCAGAGTAATTTGGCAACTCCATTTCGTGAGTGATAA 2493
713 ptle...AspGlnValArgProValIlePheLeuGluGlyGlnValP 729
: : : : : ||||| : : : : : ||||| : : : : : |||||
2494 GAGTCAAAATACAGTCCGAAACCTCCCTATTCTGTGATCGCGCAAGTGT 2543
729 heAspThrThr...GlyMetAspLysGlnThrIleAspLeuThrGlnPhe 744
: : : : : ||||| : : : : : ||||| : : : : : |||||
2544 GGAATACTAAATGGAATCAGATAAAATGGATTAGAGACCAATGCTGTG 2593
745 LysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgPr 761
|||||
2594 AAAGAC.....TTACC 2604
761 oLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaT 778
| : : : : : ||||| : : : : : ||||| : : : : : |||||
2605 CGATTATGCAATCAAGCCGCTACTCGCCCTCTACAGGTGTCGGATTCC 2654
778 hrTriPtyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeu 794
||| : : : : : ||||| : : : : : ||||| : : : : : |||||
2655 AATGCCAATCTCTAGTGGACAGCGGTATTTCTATGCTAAACCAAT 2704
795 AsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGlnIleGlySe 811
: : : : : ||||| : : : : : ||||| : : : : : |||||
2705 AAAAAATATGAAATGATGATCGAAGAGTCCCAATTTAGTATTGGGGG 2754
811 rValPhe 813
|||
2755 TTCTTTC 2761
```

seq\_name: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:US-09-135-166-7

```
seq_documentation_block:
; Sequence 7, Application US/09135166
; Patent No. 6083743
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/135,166
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
```

```
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2974 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 386..2761
; US-09-135-166-7

alignment_scores:
    Quality: 1042.00      Length: 819
    Ratio: 1.940          Gaps: 20
Percent Similarity: 65.568 Percent Identity: 31.380

alignment_block:
US-09-701-711-2 x US-09-135-166-7 ..
Align seg 1/1 to: US-09-135-166-7 from: 1 to: 2974

23 SerThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIleThrIleTh 39
425 ACAAGGACTGTGTGTTGGCCGACCTTTTGTGGCAAAAGATATTTCGTGGA 474
39 rGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPheA 56
||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
475 TGGTGTTCAGGTGACTTAGACACACAAATCCGACGAGTTTACCTGTTTC 524
56 rGluGlyGlnValValSerGluAsnGlnLeuAlaAspGlyValLysAla 72
|| ||||| ||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
525 GTGCCGTCAGCGTGTGACTGACAAATGATGTGCTAAATATTGTCGCTCT 574
73 LeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisGlnGluG 89
||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
575 TTATTCGTAAGTGGTTCGATTCGATGATGAAAGCG...CATCAAGAAGG 621
89 yArgIle...IleTyrGlnValThrGluArgProLeuIleAlaGluIleA 105
| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
622 CGATGTGCTTGTGTTAGCGTGTGCTAAATCGATCATTTTCAGATGTTA 671
105 snPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluGlyLeu 121
: : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
672 AAATCAAGGTAACCTCTATTATTCCACCTGAAGCAGCTATCAAGTAGTGCCTATA 721
122 LysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrValG 138
||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
722 GATGCTAACGGGTTTAAAGTTGGCGATATTTAATTCGAGAAAAATATA 771
138 nMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyrTyrA 155
: : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
772 TGAATTTGCCCAAAGTGTAAAGAGACACTATCAAGTAGTGCCTATA 821
155 snThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgValLys 171
||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
822 ACGCAACCGTTGAACCTATTGTCAATACGCTACCAAAATAATCGCGTGAA 871
172 LeuAspMetThrPheAlaGlyLysProAlaArgValValAspIleAs 188
: : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
872 ATTTTAATTCAAATCAATCAAGATGATAAGCCAAATTCGATCATTAAC 921
188 nIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspValLeuA 205
: : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
922 TTTCAGGGGGAACGAATCTGTTAGTAGCAGTACATTACAAGAACAAATGG 971
205 laIleLysAspAsnLysIleAsnProLeuSerLysAlaAspArgTyrThr 221
: : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
972 AATTACAACCTGATTCTTGTGGTGAATTA...TGGGGAATAAATTGAA 1018
222 GlnGluLysLeuValThrSerLeuGluAsnLeuArgAlaLysTyrLeuAs 238
```

```

1019 GGTGCGCAATTCGAGAAGATTTGACGCAATTCGTGATTATTATTTAAA 1068
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
238 nAlaGlyPheValArgPheGluIleLysAspAlaLysLeuAsnIleAsnG 255
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1069 TAATGGCTATGCCAAGACACAAATCATAAAGCGGATGTTCAAGTAAATG 1118
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
255 luAspLysAsnArgIlePheValGluIleSerLeuHisGluGlyGluGln 271
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
1119 ATGNAACAAAGTAATGTAACCATTCATGTAATGAAGGTTTACAG 1168
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
272 TyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAl 288
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1169 TATCACCTTCGTAGTCACGCATATAGTAATCTGGGAGTATGCTGC 1218
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
288 aGluLeuGluAlaLeuLys...PheLysAlaGluGluGlyPheSerG 304
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1219 CGAGCTTGAACCTTTACTTTCAGCATTTACATTTAAATGATCTTCCGCC 1268
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
304 lAlaMetLeuGluGlnThrAsnAsnIleSerThrLysPheGlyAsp 320
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
1269 GTAGTGATATGTCAGATGTAGAAAATCAATTAAGCAAACTTGGGAA 1318
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
321 AspGlyTyrTyrTyrAlaGlnIleArgProValThrArgIleAsnAspG 337
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1319 CGAGGTTACGGTAACACACAGTAATCTGTACCTCATTTTGACGATGC 1368
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
337 uSerArgThrValAspValGluTyrTyrIleAspProValHisProValT 354
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
1369 AAATAAAACATATACGGTAACTTTGTTGTGATGCTGGACGCTTTAA 1418
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
354 yrValArgArgIleAsnPheThrGlyAsnPheLysThrGlnAspGluVal 370
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1419 CTGTTCCCAACTTCGTTTGAAGGAATACCGTTTCTGCTGATGATGACT 1468
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
371 LeuArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLy 387
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1469 TTACGTCAGGAATGCGCCACAAAGAAAGAACTTGGTATTAATCACAA 1518
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
387 sIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePheLysHisV 404
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
1519 AGTGGAGTTAGGAATAATTCGCTAGATGCTACAGGTTTCTTCGAA... 1564
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
404 alThrValAspThrArg.....ProValProAsnSerProAspGlnVal 418
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1565 ..ACAGTTGAAACCGAATTCATCCTCATCAATGGTACCAATGATGAAGTG 1612
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
419 AspValAsnPheValValGluGlnProSerGlySerThrIleAl 435
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1613 GATGTCGTATATAAGTCAAGAAACGTAACACGGGTAGTATCAACTTTGG 1662
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
435 aAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSerG 452
      : |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1663 TATTTGGTTACGGTACAGAGAGTGGTATTAGTTATCAACGAAAGTGTCAA 1712
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
452 lAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1713 AGATAATTTCTTGGNAACAGGGCGGCAGTAGTATAGTGGTACGAAA 1762
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
469 SerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheTh 485
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
1763 AATGATTATGTCAGAGTGTCAATTTGGGTATACCGAGCCCTATTTTAC 1812
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
485 rValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLysT 502
      | ::::: ::::: ::::: ::::: ::::: ::::: :::::
1813 TAAAGATGGGTGA.....AGTCTTGGTGGAAATGTTTTCTTTGAAAACT 1856
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
502 yrAspAsnLys.....AsnIleSerAsnTyrValLeuAspSerTyr 515
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1857 ACGATAACTCTAAAGTGATACATCTCTAACTATAAGCGTACGACTAT 1906
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
516 GlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSe 532
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

```

1307 GGAAGTAATGTTACTTAGGTTTCCCTGTAATAAGAAATAACTCCTATTA 1956
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
532 rPheGlyLeu.....AsnAlaAspAsnThrLysLeuHisG 544
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1957 TGTAGGATTAGGCCATACCTATAATAAATAGTAACCTTTCCTCTAGAA 2006
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
544 lyGlyArgPheMetGlyIleSerAsnValLysGlnLeuMetAlaAspGly 560
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
2007 ATACCGTAAATTTATATATATCAATCAATGAAA..... 2038
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
561 GlyLysIleGlnValAspAsnAsnGlyIleProAspPheLysHisAspTy 577
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
2039 .....TTTAAAGGTAATGGCATTTAAACAAATGACTTTGATTT 2076
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
577 rThrThrTyrAsnAlaIleLeuGlyTyrAsnTyrSerSerLeuAspArgP 594
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
2077 TTCT.....TTTGGTTGGACTATAACAGCCTTAATAGAG 2111
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
594 roValPheProThrGlnGlyMetSerHisSerValAsp.....LeuThr 608
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
2112 GCTATTTCCCACTAAAGGGTTAAAGCAAGTCTTGGTGGACGAGTTACA 2161
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
609 ValGlyPheGlyAspLysThrHisGlnLysValValTyrGlnGlyAsnI 625
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
2162 ATTCAGGTTCTGATAACAAATACTACAAACTAAGTGCAGATGTACAGGG 2211
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
625 eTyrArgProPhe.....IleLysLysSerValL 635
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
2212 TTCTTACCCATTTAGACAGAGATCACCTCTGGGTTGTATCTGCAAAAGCAT 2261
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
635 euArgGlyTyrAlaLysLeuGlyTyrGlyAsn...AsnLeuProPheTyr 650
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
2262 CTGCAGGATATGCAAT...GGTTTGGAAACAACGCTTTACCGTTCTAT 2308
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
651 GluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
2309 CAAACTTATACAGCGGTGGCATTTGTTTCATTACGCGGTTTTGCTTATGG 2358
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
2359 TAGCATTTGGGCTTAACGCA.....ATTATCAAGGTCAAA 2393
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
684 lNThrThrLeu.....GlyGluValValGlyGlyAsnAlaLeu 696
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
2394 ATAATAATTTAATAAGATAGTTCTGATGTGATTTGGTGAATGCAATC 2443
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
697 AlaThrPheGlySerGluLeuIleLeuProLeuProPheLysGlyAspTr 713
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
2444 GCTACAGCTAGCGCAGAGTTAATTTGTGCCAACTCCATTTGTGAGTGATA 2493
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
713 pIle....AspGlnValArgProValIlePheIleGluGlyGlnValP 729
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
2494 GAGTCAAAATACAGTCCGAACCTCCCTATTGTTGATGCGGCAAGTGTT 2543
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
729 heAspThrThr...GlyMetAspLysGlnThrIleAspLeuThrGlnPhe 744
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
2544 GGAATACTAAATGCAATCAGATAAAATGATGATGAGACGAATGCTTG 2593
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
745 LysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgPr 761
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
2594 AAAGAC.....TTACC 2604
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
761 oLeuLeuThrClnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaT 778
      | ::::: ::::: ::::: ::::: ::::: ::::: :::::
2605 CGATTATGGCAATCAAGCGGTACTCGCGCTCTACAGGTCTCGGATTC 2654
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
778 hrTpyTyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeu 794
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
2655 AATGGCAATCTCTAGTGGACCACTGGTATTTCTTATGCTAAACCAATT 2704
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
795 AsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGlnIleGlySe 811
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
2705 AAAAAATATGAAATGATGATGTCGAACAGTTCCAATTTAGTATTGGGG 2754
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```





```
1369 AAATAAAACATTAGCGATAACCTTTGTGTGATGCTGGACGAGTTTAA 1418
      ::::::::::: :: :: :: :: :: :: :: :: :: :: ::
354 yrValargArgIleAsnPhThrGlyAsnPhelysthrGlnAspGluVal 370
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1419 CTGTTCAACCACTTCGCTTTGAAGAAATACCCGTTTCTGCTAGTAGTACT 1468
      ::::::::::: :: :: :: :: :: :: :: :: :: :: ::
371 LeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLys 387
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1469 TTAGCTGAGGAAATGCGCCAAAGAGGAACCTTGGTATAATATCACAATT 1518
      ::::::::::: :: :: :: :: :: :: :: :: :: :: ::
387 sIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePheLysHisV 404
      ::::::::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
1519 AGTTGAGTTAGGAAAAATTCGCTTAGATCGTACAGGTTTCTTCGAA... 1564
      ::::::::::: :: :: :: :: :: :: :: :: :: :: ::
404 aIThrValAspThrArg.....ProValProAsnSerProAspGlnVal 418
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1565 ..ACAGTTGAAAACCGAAATGTGATTCCTATCAATGGTAGCAATGAAGTG 1612
      ::::::::::: :: :: :: :: :: :: :: :: :: :: ::
419 AspValAsnPhValValGluGluGlnProSerGlySerSerThrIleAl 435
      ||| ||| :: :: :: :: :: :: :: :: :: :: :: :: ::
1613 GATGTCGTATATAAGTCAAGAACGTACACGGGTAGTAGTATCAACTTTGG 1662
      ||| ||| :: :: :: :: :: :: :: :: :: :: :: :: ::
435 aAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSerG 452
      : ||| ||| :: :: :: :: :: :: :: :: :: :: :: ::
1663 TATTGGTTACGGTACAGAGAGTGCTATTAGTTATCAAGCAAGTGTCAAAC 1712
      ::::::::::: :: :: :: :: :: :: :: :: :: :: ::
452 lnAsnAsnPhMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468
      : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1713 AGATATAATTTCTGGGAACAGGGCGCGCATAGTATAGTGTGTCAGGAA 1762
      ::::::::::: :: :: :: :: :: :: :: :: :: :: ::
469 SerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheTh 485
      ::::::::::: :: :: :: :: :: :: :: :: :: :: ::
1763 AATGATTATGTTACGAGTGTCAATTTGGGTATATACCGAGCCCTATTTTAC 1812
      ::::::::::: :: :: :: :: :: :: :: :: :: :: ::
485 rValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLysL 502
      : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1813 TAAAGATGGTGTA....AGTCTGGTGGAATGTTTCTTTGAAAACT 1856
      ::::::::::: :: :: :: :: :: :: :: :: :: :: ::
502 yrAspAsnLys.....AsnIleSerAsnTyrValLeuAspSerTyr 515
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1857 ACATAACTCTAAAGTGATACATCTCTAACTATAAGCGTACGAGCTTAT 1906
      ::::::::::: :: :: :: :: :: :: :: :: :: :: ::
516 GlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSe 532
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1907 GGAAGTAATGTTACTTTAGTTTCCCTGTTAAATGAAATAACTCTTATTA 1956
      ::::::::::: :: :: :: :: :: :: :: :: :: :: ::
532 rPheGlyLeu.....AsnAlaAspAsnThrLysLeuHisG 544
      ||| ||| :: :: :: :: :: :: :: :: :: :: ::
1957 TGTAGGATTAGGCATACCTATATAATAATAATAGTAATTTGCTCTAGAAT 2006
      ::::::::::: :: :: :: :: :: :: :: :: :: :: ::
544 lyGlyArgPheMetGlyIleSerAsnValLysGlnLeuMetAlaAspGly 560
      ::::::::::: :: :: :: :: :: :: :: :: :: :: ::
2007 ATACCGTAATTATATATATTCATCAATCAATGAA..... 2038
      ::::::::::: :: :: :: :: :: :: :: :: :: :: ::
561 GlyLysIleGlnValAspAsnAsnGlyIleProAspPheLysHisAspTy 577
      ::::::::::: :: :: :: :: :: :: :: :: :: :: ::
2039 .....TTTAAAGGTAATGCAATTAACCAACAAATGACTTTGATT 2076
      ::::::::::: :: :: :: :: :: :: :: :: :: :: ::
577 rThrThrTyrAsnAlaIleLeuGlyTyrAsnTyrSerSerLeuAspArgP 594
      ::::::::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
2077 TTCT.....TTTGGTTGGAACATATAACAGCCTTAATAGAG 2111
      ::::::::::: :: :: :: :: :: :: :: :: :: :: ::
594 roValPheProThrGlnGlyMetSerHisSerValAsp.....LeuThr 608
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2112 GCTATTTCCCACTAAAGGGGTTAAAGCAAGTCTTGGTGACGAGTTACA 2161
      ::::::::::: :: :: :: :: :: :: :: :: :: :: ::
609 ValGlyPheGlyAspLysThrHisGlnLysValValTyrGlnGlyAsnIl 625
      :: :: :: :: :: :: :: :: :: :: :: :: ::
2162 ATCCAGGTTCTGATAACAATACTACAACTAAGTGCAGATGTACAGG 2211
      :: :: :: :: :: :: :: :: :: :: :: :: ::
625 eTyrArgProPhe.....IleLysLysSerVal 635
      :: :: :: :: :: :: :: :: :: :: :: :: ::
```

```
2212 TTTTACCCATTAGACAGAGATCACCTCTGGTGTATCTGCAAAAGCAT 2261
      ::::::::::: :: :: :: :: :: :: :: :: :: :: ::
635 euArgGlyTyrAlaLysLeuGlyTyrGlyAsn...AsnLeuProPheTyr 650
      ||| ||| ||| ||| :: :: :: :: :: :: :: :: :: ::
2262 CTCAGGATATGCAAA...GGTTTGGAAACAAGCGTTTACCCTTCTAT 2308
      ::::::::::: :: :: :: :: :: :: :: :: :: :: ::
651 GluAsnPhThrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667
      ::::::::::: ||| ||| ||| ||| ||| ||| ||| |||
2309 CAACTTATACAGGGGTGCATGTTGTTCAATTACCGCGTTTTCCTTATGG 2358
      :: :: :: :: :: :: :: :: :: :: :: :: ::
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
      :: :: :: :: :: :: :: :: :: :: :: :: ::
2359 TAGCATTTGGCCCTAACGCA.....ATTATCAAGGTCAAA 2393
      ::::::::::: :: :: :: :: :: :: :: :: :: :: ::
684 lnThrThrLeu.....GlyGluValValGlyGlyAsnAlaLeu 696
      :: :: :: :: :: :: :: :: :: :: :: :: ::
2394 ATAATAAATTTAATAAGATAAGTTCTGATGTGTTGGTGAATGCAATC 2443
      ::::::::::: :: :: :: :: :: :: :: :: :: :: ::
697 AlaThrPheGlySerGluLeuIleLeuProLeuProPheLysGlyAspTr 713
      ||| ||| :: :: :: :: :: :: :: :: :: :: ::
2444 GCTACAGCTAGCCGACAGTTAATTGTCCTCACTCCATTTGTGAGTGATA 2493
      :: :: :: :: :: :: :: :: :: :: :: :: ::
713 pile...AspGlnValArgProValIlePheIleGluGlyGlnValP 729
      :: :: :: :: :: :: :: :: :: :: :: :: ::
2494 GAGTCAAAATACAGTCGAAACCTCCCTATTATTGTTGATCGCGCAAGTGT 2543
      :: :: :: :: :: :: :: :: :: :: :: :: ::
729 heAspThrThr...GlyMetAspLysGlnThrIleAspLeuThrGlnPhe 744
      :: :: :: :: :: :: :: :: :: :: :: :: ::
2544 GGAATACTAAATGGAATCAGATAAAATGGATTAGAGACCAATGTCTTG 2593
      :: :: :: :: :: :: :: :: :: :: :: :: ::
745 LysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgPr 761
      ||| ||| :: :: :: :: :: :: :: :: :: :: ::
2594 AAAGAC.....TTACC 2604
      ::::::::::: :: :: :: :: :: :: :: :: :: :: ::
761 oLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaT 778
      : :: :: :: :: :: :: :: :: :: :: :: ::
2605 CGATTATGCAAAATCAAGCGCTACTCGCGCTCTACAGGTGTCGGAATCC 2654
      :: :: :: :: :: :: :: :: :: :: :: :: ::
778 hrTrpTyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeu 794
      ||| :: :: :: :: :: :: :: :: :: :: :: ::
2655 AATGGCAATCTCTAGTGGACCACTGTTTCTTATGCTAAACCAATT 2704
      :: :: :: :: :: :: :: :: :: :: :: :: ::
795 AsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGlnIleGlySe 811
      :: :: :: :: :: :: :: :: :: :: :: :: ::
2705 AAAAAATACAAATGATGATGTCGAACACTTCCAATTTAGTATTGGGG 2754
      :: :: :: :: :: :: :: :: :: :: :: :: ::
811 rValPhe 813
      : |||
2755 TTCTTTC 2761
      ::::::::::: :: :: :: :: :: :: :: :: :: :: ::
seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-119-125A-3
seq_documentation_block:
; Sequence 3, Application US/08119125A
; Patent No. 5610011
; GENERAL INFORMATION:
; APPLICANT: SMITH, Hilda Elizabeth
; APPLICANT: VECCH, Uri
; TITLE OF INVENTION: DNA Sequences which code for Virulence
; TITLE OF INVENTION: Characteristics of Streptococcus suis and parts thereof, po
; TITLE OF INVENTION: Antibodies derived therefrom and the use thereof for the di
; TITLE OF INVENTION: protection against infection by S. suis in mammals, includi
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Centraal Diergeneeskundig Instituut
; STREET: Edelhertweg 15
; CITY: PH Lelystad
; STATE:
; COUNTRY: The Netherlands
; ZIP: NL-8219
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; COMPUTER: IBM compatible
```

[illegible]

```

alignment_scores:      Length: 1048
                        Quality: 154.50
                        Ratio: 0.354
                        Gaps: 55
                        Percent similarity: 41.698
                        Percent Identity: 18.989

alignment_block:
  HS-09-701-711-2   x   US-08-119-125A-3   ..

```









309 nThrThrAsnAsnIleSerThrLys.....P 318  
 : : : : :  
 1479 .TCAACCATCAATGTATCAGGCAAGAAAAAGCGGCGCTATTGTGT 1527  
 : : : : :  
 318 heGlyAsp.....AspGlyTyrTyrTyrAlaGlnIleArgPro 330  
 : : : : :  
 1528 GGGCGGATATTGGCTTAATTTGACGGCAATATTAAACGCTCAA..... 1568  
 : : : : :  
 331 ValThrArgIleAsnAspGluSerArgThr..... 340  
 : : : : :  
 1569 .....GGTAGTGGTATATCGCTAAACCGGTGGTGGTGGAGACGCTC 1612  
 : : : : :  
 341 .....ValAspVal.....GluT 345  
 : : : : :  
 1613 GGGGCATGATTATTTCATCAAGACAATGCAATTTGTAGCGCCAAAGAGT 1662  
 : : : : :  
 345 yrTyrIleAspPro.....ValHisProValTyrValArgArg 357  
 : : : : :  
 1663 GGTGTGTAGCCGGTAATATGTATCTATTAAATGCAGAAACAGCAGACGC 1712  
 : : : : :  
 358 IleAsn.....PheThrGlyAsnPheLysThrGlnAs 368  
 : : : : :  
 1713 AGCAATACCTTCAGAACGAGTAAACACGGGATCCGGGAATAGTGCCAG 1762  
 : : : : :  
 368 pGluValLeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerA 385  
 : : : : :  
 1763 CACCCCAACGAACAAA.....GAAAGACACACATTAAACAA 1800  
 : : : : :  
 385 snGlnLysIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePhe 401  
 : : : : :  
 1801 ACACAACCTCTTGAG.....AGTATACTAAAAAAGGTACCTTT 1838  
 : : : : :  
 402 LysHisValThrValAspThrArgProValProAsnSerProAspGlnVa 418  
 : : : : :  
 1839 GTTAACATCCTGCTAATCAACGCACTATGTCATAGC..... 1877  
 : : : : :  
 418 lAspValAsnPheValValGluGlnProSerGlySerSerThrIleA 435  
 : : : : :  
 1878 .TCCATTAAATTA.....TCCAATGGCAGCTTAACCTCTTT 1911  
 : : : : :  
 435 lAlaGlyTyrSerGlnSerGlyValThrPheGlnPheAspValSer 451  
 : : : : :  
 1912 GGAGTCAGGCGCGGCGGTGAGATT..... 1946  
 : : : : :  
 452 GlnAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerA 468  
 : : : : :  
 1947 ...AACACGATATTACCCGCGGTGATGAT.....ACCAG 1978  
 : : : : :  
 468 gSerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheT 485  
 : : : : :  
 1979 AGGTGCAAACTTAACAATTTACTCAGCGCGGTGGTTGATGTTCAATAAA 2028  
 : : : : :  
 485 hrValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLys 501  
 : : : : :  
 2029 ATATCTCACTCGGGCGGCAAGGTAAACATAACATTACAGCTAAACAAGAT 2078  
 : : : : :  
 502 TyrAspAsnLysAsnIleSerAsnTyrValLeuAspSerTyrGlyLys 518  
 : : : : :  
 2079 ATCGCCTTTGAGAAAGGAAGAACCAAGTCATTACAGGTCAA...GGGAC 2125  
 : : : : :  
 518 rLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSerPheGlyL 535  
 : : : : :  
 2126 TATTACCTCAGGC.....AATCAAAA.....GGT 2151  
 : : : : :  
 535 euAsnAlaAspAsnThrLysLysLysGly..... 544  
 : : : : :  
 2152 TTAGATTTAATAATGCTCTCTAAACGGCACTGGCAGCGGACTGCAATTC 2201  
 : : : : :  
 545 .....GlyArgPheMetGlyI 550  
 : : : : :  
 2202 ACCACTAAAGAACCAATAAATACGCTATCACAAATAAATTTGAAGGGAC 2251

550 eSerAsnValLysGlnLeuMetAlaAspGlyGlyLysIleGlnValAspA 567  
 : : : : :  
 2252 TTTAAATATT.....TCAGGAAAGTGAACATC...T 2280  
 : : : : :  
 567 snAsnGlyIleProAspPheLysHisAspTyrThrThrTyrAsnAlaIle 583  
 : : : : :  
 2281 CAATGGTTTACCTAAATAAAGTGAATGATATAAATTCAAAGGACGC 2330  
 : : : : :  
 584 LeuGlyTyrAsnTyrSerSerLeuAsp..... 592  
 : : : : :  
 2331 ACTTACTGGAAATTAACCTCCTTAATGTTCCGAGAGTGGCGAGTTAA 2380  
 : : : : :  
 593 .....ArgP 594  
 : : : : :  
 2381 CCTCACTATTGACTCCAGAGGAAGCATAGTCAGGCGACACTTACCCAGC 2430  
 : : : : :  
 594 roValPheProThrGlnGlyMetSerHisSerValAspLeuThrValGly 610  
 : : : : :  
 2431 CT...TATATATTAAACGGTATATCATTCACAAAGACACTACC..... 2471  
 : : : : :  
 611 PheGlyAspLysThrHisGlnLysValValTyrGlnGlyAsnIleTyrAr 627  
 : : : : :  
 2472 TTTAATGTTGAACGAATGCAAGAGTCAACTTTGAC..... 2507  
 : : : : :  
 627 gProPheIleLysLysSerVal...LeuArgGlyTyrAlaLysLeuGlyT 643  
 : : : : :  
 2508 .....ATCAAGGCACCAATAGGATAAATAAGTATTCTAGTTGAAT 2550  
 : : : : :  
 643 yrGlyAsnAsnLeuProPheTyrGluAsnPheTyrAlaGlyGlyTyrGly 659  
 : : : : :  
 2551 ACGCATCA.....TTTAATGGAAACATTTTCAGTTTCGGGAGGGGG 2591  
 : : : : :  
 660 SerValArgGlyTyrAspGlnSerSerLeuGlyProArgSerGlnAla... 675  
 : : : : :  
 2592 AGTGT.....GATTTCACTTCGCTCATCTCAACGTCCTCA 2632  
 : : : : :  
 676 .....TyrLeuThrAlaArgGlyG 683  
 : : : : :  
 2633 AACCCCGGTGTAGTTATAAATCTTAATACTTTAATGTTTCAACAGGT 2682  
 : : : : :  
 683 lndln.....ThrThrLeuGlyGluValValGlyGlyAsnAlaLeu 696  
 : : : : :  
 2683 CAAGTTTAAGATTAAACCTTCAGGCTCAACAAAACTGGCTTCTCAATA 2732  
 : : : : :  
 697 .....AlaThrPheGlySerGluLeuLeuPr 706  
 : : : : :  
 2733 GAGAAAGATTTAATTTAATGCCCGGAGGCAACATAACACTTTTGCA 2782  
 : : : : :  
 706 oLeuProPheLysGlyAspTyrIleAspGln.....ValArgProV 720  
 : : : : :  
 2783 AGTTGAAGGCACCGATGGAATGTTGTTAAAGGCATTGTAGCCAAAAAAA 2832  
 : : : : :  
 720 alllePheIleGluGlyGlyGlnValPheAspThrThrGlyMetAspLys 736  
 : : : : :  
 2833 ACATAACCTTTGAAGGAGTAAACATC.....ACCTTTGGCTCCAGGAAA 2876  
 : : : : :  
 737 GlnThrIleAspLeuThrGlnPheLysAspProGlnAlaThrAlaGluG 753  
 : : : : :  
 2877 GCGTAACAGAAATC.....GAAGGCAATGTTACTATCAATAA 2914  
 : : : : :  
 753 nAsnAlaLysAlaAla.....AsnArgP 761  
 : : : : :  
 2915 CAACGCTAACGTCACCTTATCGTTTCGGATTTTGACAAACCAACAAAAC 2964  
 : : : : :  
 761 roLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAla 777  
 : : : : :  
 2965 CTTTAACTATTAAAAAAGATGTCATCATTAATAGCGCAACCTTACCGCT 3014  
 : : : : :  
 778 ThrTyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLe 794  
 : : : : :  
 3015 .....GGAGCAATATTGTCAATATAGCCGGAATCT 3046  
 : : : : :  
 794 uAsnLysLysGlnAsnAspGlnThrAspThrVal.....GlnPheG 808

731 TGTTACATCTAACCAAAATCTCCCAATTAAGAGGATTTTAGATTCTAAC. 779  
57 euGlyGlnVal.....ValSerGluAsnGlnLeuAlaAspGlyVallys 71  
780 ..GGACAAGTCTTTTAAATCAACCAAAATGATATACCAATAGGTAAAGAC 827  
72 AlaLeuTyrAlaThrGlyAsnPhSer.....AspValGlnVa 84  
828 GCAATTATTAACTAATGGCTTTACGGCTTCTAGCTAGACATTCTCTAA 877  
84 lTyrHisGlnGluGlyArg...IleIleTyrGlnValThrGluArgProL 100  
878 CGAAACATCAAGCGCGTAATTTTACCTTCGAGCAAAACCAAGATAAG 927  
100 euIleAlaGluIleAsnPhGluGlyAsnArgLeuIleProLysGluGly 116  
928 CGCTCGCTGAAATGTGAATCAGCGTTTAAATCTGCTCGTAAAGACGGC 977  
117 .....LeuGlnGluGlyLeuLysAsnAlaGly..... 125  
978 AGTGTAATCTTATTGGTGGCAAGTGAAAAACGAGGCTGCTAGTAGCGT 1027  
126 .....LeuAlaValGlyGlnProLeuLysGlnAlaT 136  
1028 AAATGGTGGCAGCATTTCTTTACTCGCAGGGCAAAATC.....A 1068  
136 hrValGlnMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGly 152  
1069 CCATCAGCGATATA..... 1082  
153 TyrTyrAsnThrGluIleThrVallysGlnThrMetLeuAspGlyAsnAr 169  
1083 ...ATAAACCCCAACCATTTACTTACAGCATTCGCCGCGCTGAAATGAAGC 1129  
169 gValLysLeuAspMetThrPheAlaGluGlyLysProAlaArgValVala 186  
1130 GGTCAATCTGGCGGATATTTTGCCAAAGCGCGT.....A 1164  
186 spIleAsnIle.....IleGlyAsnGlnHisPheSerAspAla 198  
1165 ACATTAATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1214  
199 AspLeuIleAspValLeuAlaIleLysAspAsnLysIleAsn...ProLe 214  
1215 GATTCGTGAAGC.....AAAGATAAAAGCGCGCAATATTTCTTCT 1252  
214 uSerLysAlaAspArgTyrThrGlnGlyLysLeuValThrSerLeuGlu 231  
1253 TTCGCCCAAGAGGGTGAAGCGGAATTCGCGTGTAATTCGCGCTCAA 1302  
231 snLeuArgAlaLysTyrLeuAsnAlaGlyPheValArgPheGluIleLys 247  
1303 ATCAGCAAGCT.....AAA 1316  
248 AspAlaLysLeuAsnIleAsnGluAspLysAsnArgile..... 260  
1317 GCGCGCAAGCTGATGATTACAGCGGATAAGTCACATTTAAACACAGGTGC 1366  
261 .PheValGluIleSerLeuHisGlu...GlyGluGlnTyrArgPheGlyG 276  
1367 AGTTATCGACCTTTCAGGTAAGAGAGGGGAGAACTTACCTTGGCGGTG 1416  
276 InThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAlaGluLeuGluAla 292  
1417 ACGAGCGCGCGAAGGT.....AAAAAGGCATTCATTA 1451  
293 LeuLeuLysPheLysAlaGluGluGlyPheSerGlnAlaMetLeuGluG 309  
1452 GCAAGAAACCTCTTTAGAAAAAGGC..... 1478  
309 nThrThrAsnAsnIleSerThrLys.....P 318  
1479 .TCAACCATCAATGTATCAGGCAAGAAAAAGGCGGCGCTATTGTGT 1527

3047 TACGGTTGAAGTAACGCTATTTCAAGCTATCAAAATTTCACTTTTA 3096  
808 InIleGlySerValPhe 813  
3097 ATGTAGCGGCTGTTT 3113  
seq\_name: /cgn2.6/ptodata/2/lna/5A\_comb.seq:US-08-302-832-1  
seq documentation\_block:  
; Sequence 1, Application US/08302832  
; Patent No. 5603938  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen J  
; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd.  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; STREET: Bldg. 1  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/302,832  
; FILING DATE: 16-SEP-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9205704.1  
; FILING DATE: 16-MAR-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US pct/us93/02166  
; FILING DATE: 16-MAR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Berkstreser, Jerry W  
; REGISTRATION NUMBER: 22,651  
; REFERENCE/DOCKET NUMBER: 1038-404  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 415-0810  
; TELEFAX: (703) 415-0813  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5116 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-08-302-832-1  
alignment\_scores:  
Quality: 147.00 Length: 939  
Ratio: 0.347 Gaps: 53  
Percent Similarity: 45.154 Percent Identity: 20.554  
alignment\_block:  
US-09-701-711-2 x US-08-302-832-1 ...  
Align seg 1/1 to: US-08-302-832-1 from: 1 to: 5116  
27 GlnAlaAlaAspPheMetAlaAsnAspIleThrIleThrGlyLeuGlnAr 43  
681 GAAATGTCAGTGTTCACAGAAACAACTCCGCGCTATTCACCG 730  
43 gValThr.....IleGluSerLeuGlnSerValLeuProPheArgL 57  
|||||||



```

318 heGlyAsp.....AspGlyTyrTyrTyrAlaGlnIleArgPro 330
:|||||
1528 GGGCGGATATTGGTTAAATGACGGCAATTAACGCTCAA.....1568
:|||||

331 valThrArgIleAsnAspGluSerArgThr.....340
:|||||
1569 .....GGTAGTGGTGATATCGCTAAACCGGGTGTGTGGAGACGTC 1612
:|||||

341 .....ValAspVal.....GluT 345
:|||||

1613 GGGGCGATGATTTATTCATCAAGACAATGCAATTTGTGACGCCAAGAGT 1662
:|||||

345 yrTyrIleAspPro.....ValHisProValTyrValArgAq 357
:|||||

1663 GGTGTGTAGCCCGGATATGATCTATTAAATGCAGAAACAGCAGACGC 1712
:|||||

358 IleAsn.....PheThrGlyAsnPheLysThrGlnAs 368
|||
1713 AGCAATACTTCAGAGACGATGAATACACGGGATCCGGGAATAGTGCCAG 1762
:|||||

368 pGluValLeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerA 385
:|||||
1763 CACCCCAAAACCAACANA.....GAAAAGACAAACATTAAACA 1800
:|||||

385 snGlnLysIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePhe 401
|||
1801 ACACAACTCTTGAG.....AGTATACTAAAAAAGTACCTTT 1838
:|||||

402 LysHisValThrValAspThrArgProValProAsnSerProAspGlnVa 418
:|||||
1839 GTTAACATCACTGCTAATCAACGCATCTATGTCATAGC.....1877
:|||||

418 laspValAsnPheValValGluGlnProSerGlySerSerThrIleA 435
:|||||
1878 .TCCATTAATTA.....TCCAAATGGCAGCTTAACCTTT 1911
:|||||

435 laAlaGlyTyrSerGlnSerGlyValThrPheGlnPheAspValSer 451
:|||||
1912 GGAGTCAGGTCGGACGGTGGCGGTGAGATT.....1946
:|||||

452 GlnAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerAr 468
:|||||
1947 ..AACACGATATTACCACCGGTGATGAT.....ACCAG 1978
:|||||

468 gSerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPhe 485
:|||||
1979 AGGTGCAAACTTAACAATTTACTACGGCGGTGGGTGATGTTCAATAA 2028
:|||||

485 hrValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLys 501
:|||||
2029 ATATCTCACTCGGGCGCAAGGTAAACATAACATTACAGCTAAACAAGAT 2078
:|||||

502 TyrAspAsnLysAsnIleSerAsnTyrValLeuAspSerTyrGlyGlySe 518
:|||||
2079 ATCGCCTTTGAGAAAGGAAGAACCAACCAAGTCATTACAGGTCAA...GGGAC 2125
:|||||

518 rLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSerPheGlyL 535
:|||||
2126 TATTAACCTCAGGC.....AATCAAAA.....GGTT 2151
:|||||

535 euAsnAlaAspAsnThrLysLysHisGly.....544
:|||||
2152 TTAGATTTAATAATGCTCTCTAAACGGCAGCTGGCAGCGAGCTGCAATTC 2201
:|||||

545 .....GlyArgPheMetGlyI 550
:|||||
2202 ACCACTAAAGAACCAATAATACGCTATCACAAATAAATTTGAAGGGAC 2251
:|||||

550 eSerAsnValLysGlnLeuMetAlaAspGlyGlyLysIleGlnValAspA 567
:|||||
2252 TTAAATATT.....TCAGGAAAGTGAACATC...T 2280
:|||||

```

```

567 snAsnGlyIleProAspPheLysHisAspTyrThrThrTyrAsnAlaIle 583
:|||||
2281 CAATGGTTTACCTAAAAATGAAAGTGGATATGATAAATTCAAAGGACGC 2330
:|||||

584 LeuGlyTyrAsnTyrSerSerLeuAsp.....592
:|||||
2331 ACTTACTGGAATTTAACTCTCTTAAATGTTTCCGAGAGTGGCGAGTTAA 2380
:|||||

593 .....Argp 594
:|||||

2381 CTTCACTATTGACTCCAGAGGAAGCATAGTCAGGCACACTTACCACGC 2430
:|||||

594 roValPheProThrGlnGlyMetSerHisSerValAspLeuThrValGly 610
|||
2431 CT...TATAATTAAACGGGTATATCATTCACAAACACACTACC.....2471
:|||||

611 PheGlyAspLysThrHisGlnLysValValTyrGlnGlyAsnIleTyrAr 627
:|||||
2472 TTATAATGTTGACGAAATGCAAGAGTCAACTTTGAC.....2507
:|||||

627 gProPheIleLysLysSerVal...LeuArgGlyTyrAlaLysLeuGlyT 643
:|||||
2508 .....ATCAAGGCACCAATAGGATAAATAAGTATCTTAGTTTGAATT 2550
:|||||

643 yrGlyAsnAsnLeuProPheTyrGluAsnPheTyrAlaGlyGlyTyrGly 659
:|||||
2551 ACGCATCA.....TTTAAATGGAACATTTTCAGTTTCGGGAGGGGG 2591
:|||||

660 SerValArgGlyTyrAspGlnSerSerLeuGlyProArgSerGlnAla... 675
:|||||
2592 AGGTGTT.....GATTTACACTTCGCGCTCATCTCTAAGCTCA 2632
:|||||

676 .....TyrLeuThrAlaArgGlyG 683
:|||||
2633 ACCCCCGGTGAGTTATAAATCTTAATACTTTAATGTTTCAACAGGGT 2682
:|||||

683 lngln.....ThrThrLeuGlyGluValValGlyGlyAsnAlaLeu 696
:|||||
2683 CAAGTTTAAGATTAAAACTTCAGGCTCAACAAAACTGCGTCTCAATA 2732
:|||||

697 .....AlaThrPheGlySerGluLeuLeuLeuPr 706
:|||||
2733 GAGAAAGATTTAACTTTAAATGCCACCGGGAACATAACACTTTTGCA 2782
:|||||

706 oLeuProPheLysGlyAspTyrIleAspGln.....ValArgProV 720
:|||||
2783 AGTTGAGGCACCGGATGGAATGATGTTAAAGGCATTTAGCCAAAAAAA 2832
:|||||

720 alilePheIleGluGlyGlyGlnValPheAspThrThrGlyMetAspLys 736
:|||||
2833 ACATAACCTTTGAAGGAGGTAAACATC.....ACCTTTGGCTCCAGAAA 2876
:|||||

737 GlnThrIleAspLeuThrGlnPheLysAspProGlnAlaThrAlaGluG 753
:|||||
2877 GCGGTAAACAGAAATC.....GAAGCAATGTTACTATCAATAA 2914
:|||||

753 nasnAlaLysAlaAla.....AsnArgp 761
:|||||
2915 CAACGCTAACGTCACCTTATCGTTCCGATTTTGACACCACTCAAAAAC 2964
:|||||

761 roLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAla 777
:|||||
2965 CTTTAACATTAAAAAAGATGTCAATCAATTAATAGCGCAACCTTACCGCT 3014
:|||||

778 ThrTyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLe 794
:|||||
3015 .....GGAGCAATATGTCATATATAGCCGGAATCT 3046
:|||||

794 uAsnLysLysGlnAsnAspGlnThrAspThrVal.....GlnPheG 808
:|||||
3047 TACCGTTGAAGTAACGCTAATTTCAAGCTATCAACAATTTTCACCTTTA 3096
:|||||

808 lngIleGlySerValPhe 813

```

```

      :|||||:|||||:
3097 ATGTAGCGGCTGTTT 3113
seq_name: /cqn2_6/ptodata/2/ina/3b_COMB.seq:US-08-530-198-1
seq_documentation_block:
; Sequence 1, Application US/08530198
; Patent No. 5869065
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GENE III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,198
; FILING DATE: 13-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: JWB-1186
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 1:
; LENGTH: 5116 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-530-198-1

alignment_scores:
  Quality: 147.00      Length: 939
  Ratio: 0.347        Gaps: 53
  Percent Similarity: 45.154  Percent Identity: 20.554

alignment_block:
US-09-701-711-2 x US-08-530-198-1 ..
Align seg 1/1 to: US-08-530-198-1 from: 1 to: 5116
27 GlnAlaAlaAspPheMetAlaAsnAspIleThrIleThrGlyLeuGlnAr 43
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
681 GAAATCGTCGAGTCTTTTACAGAAACCAACACTCCGCGGTATTCACCG 730
43 gValThr.....IleGluSerLeuGlnSerValLeuProPheArgL 57
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
731 TGTATCATCTAACCAATCTCCCAATTAAGGGATTTCATTCCTAAC. 779
57 euGlyGlnVal.....ValSerGluAsnGlnLeuAlaaspGlyValLys 71
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
780 ..GGACAAGTCTTTTAAATCAACCAAAATGGTATCAATAGTAAGAC 827
72 AlaLeuTyraAlaThrGlyAsnPheSer.....AspValGlnVa 84
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
828 GCAATTATTAACACTAATGCGCTTACGGCTTCTACGCTAGACATTCTAA 877
```

```

84 lTyTHisGlnGluGlyArg...IleIleTyTGlnValThrGluArgProL 100
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
878 CGAAACATCAAGCGCGGTAAATTCACCTTCGAGCAAAACCAAGATAAAG 927
100 euIleAlaGluIleAsnPheGluGlyAsnArgLeuIleProLysGluGly 116
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
928 CGCTCGCTGAAATTTGTGAATCACGCTTTTAATTACTGTCGGTAAAGACGC 977
117 .....LeuGlnGluGlyLeuLysAsnAlaGly..... 125
978 AGTGTAATCTATTGTGGCGAAAGTGAATAACGAGGGTGTGATTAGCGT 1027
126 .....LeuAlaValGlyGlnProLeuLysGlnAlaT 136
1028 AAATGGTGGCAGCATTTCTTTTACTCGCAGCGGCAAAATC.....A 1068
136 hrValGlnMetIleGluThrGluLeuThrAsnGlnTyIleSerGlnGly 152
1069 CCATCAGCGGATATA..... 1082
153 TyrTyAsnThrGluIleThrValLysGlnThrMetLeuAspGlyAsnAr 169
1083 ...ATAAACCCCAACCATTTACTTACAGCATTCGCCGCCCTGAAATGAAGC 1129
169 gValLysLeuAspMetThrPheAlaGluGlyLysProAlaArgValValA 186
1130 GGTCAATCTGGCGGATATTTTTCGCAAGGCGGT.....A 1164
186 spIleAsnIle.....IleGlyAsnGlnHisPheSerAspAla 198
1165 ACATTAATGTCCGTCTGCCACTATTCGAACCAAGGTAAACTTTCTGCT 1214
199 AspLeuIleAspValLeuAlaIleLysAspAsnLysIleAsn...Prole 214
1215 GATTCGTGAAGC.....AAAGATAAAGCGGCAATATTGTTCT 1252
214 uSerLysAlaAspArgTyThrGlnGluLysLeuValThrSerLeuGluA 231
1253 TTCGCGCAAAAGAGGTGAAGCGGAAATTTGGCGGTGTAAATTTCCGCTCAA 1302
231 snLeuArgAlaLysTyThrLeuAsnAlaGlyPheValArgPheGluIleLys 247
1303 ATCAGCAAGCT.....AAA 1316
248 AspAlaLysLeuAsnIleAsnGluAspLysAsnArgIle..... 260
1317 GCGCGCAAGCTGATGATTACAGCGGATAAAGTCACATTAATAAACAGGTGC 1366
261 .PheValGluIleSerLeuHisGlu...GlyGluGlnTyArgPheGlyG 276
1367 AGTTATCGACCTTTTCAGGTAAAGAGGGGAGAACTTACCTTGGCGGTG 1416
276 lnThrGlnPheLeuGlyAsnLeuThrTyThrGlnAlaGluLeuGluAla 292
1417 ACAGCGCGCGCAAGGT.....AAAAAGGCATTCAATTA 1451
293 LeuLeuLysPheLysAlaGluGlyPheSerGlnAlaMetLeuGluG 309
1452 GCAAGAAAAACCTCTTTAGAAAAAGC..... 1478
309 nThrThrAsnAsnIleSerThrLys.....P 318
1479 .TCAACCATCAATGTATCAGGCAAGAAAAAGCGCGCGCTATTGTGT 1527
318 heGlyAsp.....AspGlyTyTyThrAlaGlnIleArgPro 330
1528 GGGCGCATATTGCGTTAATTGACGGCAATATTAAACGCTCAA..... 1568
331 ValThrArgIleAsnAspGluSerArgThr..... 340
1569 .....GGTAGTGTGATATCGCTAAACACCGGTGGTGGTGGAGACGTC 1612
```

341 .....ValAspVal...Glnr 345  
1613 GGGCGATGTTTATTCATCAAGACAAATGCAATTTGTGCGCCCAAGAGT 1662  
345 yTrIleAspPro.....ValHisProValTyrValArg 357  
::: |||||  
1663 GGTGTGTAGCCCGGATAATGTATCTATTAAATCAGAAACAGCAGGACGC 1712  
358 IleAsn.....PheThrGlyAsnPhelysthrGlnAs 368  
|||  
1713 AGCAATACCTCAGAGACGATGAATACAGGGATCCGGGAATAGTGCCAG 1762  
368 pGluValLeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerA 385  
: |||||  
1763 CACCCCAAAACGAAACAA.....GAAAGACAAACATTAAACAA 1800  
385 snGlnLysIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePhe 401  
|| :|||  
1801 ACACAACTCTGTAG.....AGTATACTAAAAAAGGTACCTTT 1838  
402 LysHisValThrValAspThrArgProValProAsnSerProAspGlnVa 418  
: |||||  
1839 GTTAACATCAGCTCTAATCAACGCATCTATGCAATAGC..... 1877  
418 lAspValAsnPheValValcIuGlnProSerGlySerSerThrIleA 435  
: |||||  
1878 .TCCATTAATTTA.....TCCAATGGCAGCTTAACCTTT 1911  
435 laAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSer 451  
::: |||||  
1912 GGAGTGAGGTCGCGCGGTGGCGGCTTGAGATT..... 1946  
452 GlnAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerAr 468  
|||  
1947 ...AACACGATATTACCGCGGTGATGAT.....ACCAG 1978  
468 gSerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheT 485  
|||  
1979 AGGTGCAAACTTAACAATTACTACAGCGCGCTGGGTGATGTTTCATAAAA 2028  
485 hrValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLys 501  
: |||||  
2029 ATATCTCAGTGGCGGCGAGGTAAACATAACATTACAGCTAAACAAAGAT 2078  
502 TyrAspAsnLysAsnIleSerAsnTyrValLeuAspSerTyrGlyLys 518  
: |||||  
2079 ATGCGCTTTGAGAAAGGACCAACCAAGTCATTACAGGTCAA...GGAC 2125  
518 rLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSerPheGlyL 535  
: |||||  
2126 TATTACCTCAGGC.....AATCAAAA.....GGTT 2151  
535 euAsnAlaAspAsnThrLysLysHisGly..... 544  
2152 TTAGATTTAATAATGTCCTCTAAACGGCACTGGCAGCGGACTGCAATTC 2201  
545 .....GlyArgPheMetGlyI 550  
2202 ACCACTAAAGAACCAATAAATACGCTATACACAATAAATTTGAAGGAC 2251  
550 eSerAsnValLysGlnLeuMetAlaAspGlyGlyLysIleGlnValAspA 567  
: |||||  
2252 TTTAAATATT.....TCAGGGAAGTGAACATC...T 2280  
567 snAsnGlyIleProAspPheLysHisAspTyrThrThrTyrAsnAlaIle 583  
: |||||  
2281 CAATGTTTACCTAAATAAGTGAAGTGGATATGATAAATTCAAAGGACGC 2330  
584 LeuGlyTriAsnTyrSerSerLeuAsp..... 592  
2331 ACTTACTGGAATTAACCTCTTAATGTTTCGAGAGTGGCGAGTTTAA 2380  
593 .....Argp 594

2381 CCTCACTATTGACTCCAGAGGAGCGATAGTCGAGGCACACTTACCACGC 2430  
594 roValPheProThrGlnGlyMetSerHisSerValAspLeuThrValGly 610  
|| :|||  
2431 CT...TATATTTAAACGGGTATATCATCTTCAACAAAGACACTACC..... 2471  
611 PheGlyAspLysThrHisGlnLysValValTyrGlnGlyAsnIleTyrAr 627  
|||  
2472 TTTAATGTTGAACGAAATGCAAGAGTCAACTTTGAC..... 2507  
627 gProPheIleLysLysSerVal...LeuArgGlyTyrAlaLysLeuGlyT 643  
|||  
2508 .....ATCAAGGCACCAATAGGATAAATAAGTATCTTAGTTGAATT 2550  
643 yrcGlyAsnAsnLeuProPheTyrGluAsnPhetYrAlaGlyGlyTyrGly 659  
|| :|||  
2551 ACCCATCA.....TTTAATGGAAACATTTTCAGTTTCGGGAGGGGG 2591  
660 SerValArgGlyTyrAspGlnSerSerLeuGlyProArgSerGlnAla... 675  
|||  
2592 AGTGT.....GATTCACACTTCGCGCTCATCTCTTAACGTCCA 2632  
676 .....TyrLeuThrAlaArgArgGlyG 683  
|||  
2633 AACCCCGGTGTAGTTATAAATCTTAAATACTTTAATGTTTCAACAGGT 2682  
683 lngln.....ThrThrLeuGlyGluValValGlyGlyAsnAlaLeu 696  
|| :|||  
2683 CAAAGTTTAAGATTAAACCTTCAGGCTCAACAAAACCTGGCTTCTCAATA 2732  
697 .....AlaThrPheGlySerGluLeuLeuPr 706  
|||  
2733 GAGAAAGATTTAACTTTAAATGCCACGGAGGCACACATAACACTTTTGA 2782  
706 oLeuProPheLysGlyAspTrpIleAspGln.....ValArgProv 720  
: |||  
2783 AGTTGAAGCAGCAGCATGGAATGATTTGTAAGGCATTGTAGCCAAAATA 2832  
720 alIlePheIleGluGlyGlnValPheAspThrThrGlyMetAspLys 736  
|| :|||  
2833 ACATAACCTTTGAAGGAGGTAAATC.....ACCTTTGGCTCCAGGAAA 2876  
737 GlnThrIleAspLeuThrGlnPheLysAspProGlnAlaThrAlaGluG 753  
: |||  
2877 GCGTAAACAGAAATC.....GAAGGCAATGTTACTATCAATAA 2914  
753 nAsnAlaLysAlaAla.....AsnArgp 761  
: |||||  
2915 CAACGCTAAGTCACCTCTTATCGGTTTCGGATTTTGACACACCATCAAAAAC 2964  
761 roLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAla 777  
|||  
2965 CTTTAACTATTAAGAAAGATGTCATCATTAATAGCGCAACCTTACCCT 3014  
778 ThrTrpTyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLe 794  
|||  
3015 .....GGAGGCAATATTGTCAATATAGCCGGAATCT 3046  
794 uAsnLysLysGlnAsnAspGlnThrAspThrVal.....GlnPheG 808  
: |||  
3047 TACCGTTGAAAGTAAGCTAATTTCAAAAGCTATCACAAATTTTCACCTTTA 3096  
808 lnIleGlySerValPhe 813  
3097 ATGTAGCGCGCTGTGT 3113

seq\_name: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:us-08-469-880-1

seq\_documentation\_block:  
; Sequence 1, Application US/08469880  
; Patent No. 5876733  
; GENERAL INFORMATION:

84 lTyrHisGlnGluclyArg...lleIleTyrcInValThrGluArgProL 100  
::: : ::::: ||| |::: :  
878 CGAAACATCAAGCGGTAAATTCACCTTCGACGAAAACCAAGAATAAAG 927  
  
100 euIIeAlaGluIleAsnPhcGluGlyAsnArqLeuIIeProLySGluGly 116  
::: ||||| ||| |::: :  
928 CGCTCGTGTAATTTGAATCAGGGTTAAATTACTGTCGGTAAAGACGCC 977  
  
117 ..... LeuGlnGluGlyLeuLySAsnAlaGly..... 125  
1978 AGTGTAAATCTTATGGTGCAAGAGTGAAACACGAGGGTGTGATTAGCGT 1027  
  
126 ..... LeuAlaValGlyGlnProLeuLySGlnAlaTf 136  
||| : ||||| |::: :  
1028 AAATGGTGGCAGCATTTCTTTACTCGCAGGCAAAAAATC.....A 1068  
  
136 hrValGlnMetIleGluThrGluLeuThrAsnGlnTyrlleSerGlnGly 152  
||||: : |||  
1069 CCATCAGCCGATATA..... 1082  
  
153 TyrTyraSnThrGluIleThrVallYsGlnThrMetLeuAspGlyAsnAr 169  
||| | ||||| : : : : :  
1083 ... ATAACCACAACCATTTACTACAGCATTCGCAGCGCTCAAATGAAGC 1129  
  
169 gVallYsLeuaspMetThrPheAlacLuGlyLysProAlaArgValVala 186  
||| : ||| | ||||| : |||  
1130 GGTCAATCTGGCGATATTTTTGCCAAAGCGGT.....A 1164  
  
186 spIIeAsnIle.....IlleGlyAsnGlnHisPheSerAspAla 198  
: ||||| : ||| | |||||  
1165 ACRTAAATGTCGTGCTGCCACTATTTCGAAACCAAGGTAAACTTCTGCT 1214  
  
199 AspLeuIleAspValLeuAlaIleLysAspAsnLySileAsn...ProLe 214  
||| : : : : : ||||| : : : : :  
1215 GATTCGTGAAGC.....AAGATATAAGCGGCAATATWTGTTCT 1252  
  
214 uSerLySAlaAspArgTyThrGlnGluLySLeuValThrSerLeuGluA 231  
TTTTCCGCAAGAGGGTGACGGGAATTCGCGGTAAATTCGCTCAAA 1302  
  
231 snLeuArgAlaLySTyrrLeuAsnAlaGlyPheValArgPheGluIleLyS 247  
||| : ||| | ||||| : |||  
1303 ATCAGCAAGCT.....AAA 1316  
  
248 AspAlaLySLeuAsnIleAsnGluAspLySAsnArgIle..... 260  
: ||||| : ||||| : ||||| :  
1317 GCGCGCAAGCTGATGATTACGCGGATTAAGTCAATTAACAAACAGGTGC 1366  
  
261 .PheValGluIleSerLeuHisGlu...GlyGluGlnTyrrArgPheGlyG 276  
: : : : : ||||| : ||| :  
1367 AGTTATCGACCTTTTCAGGTAAAGAAGGGGAGAAACTTACCTTGGCGGTG 1416  
  
276 lnThrGlnPheLeuGlyAsnLeuThrTyrrThrGlnAlaGluLeuGluAla 292  
: : : : : |||  
1417 ACGAGCGCGCGAAGGT.....AAAAGGGCAATTCATTA 1451  
  
293 LeuLeuLySPhelySAlaGluGluGlyPheSerGlnAlaMetLeuGluCl 309  
||| : ||||| : ||||| :  
1452 GCAAGAAACCTCTTTAGAAAAAGGC..... 1478  
  
309 nThrThrAsnIleSerThrLyS.....P 318  
: ||| : ||||| : ||| :  
1479 .TCAACCATCAATGATATCAGCAAGAAAGGCGGACCGCTATTGTCT 1527  
  
318 heGlyAsp.....AspGlyTyrrTyrrAlaGlnIleArgPro 330  
: ||||| : ||||| : ||||| :  
1528 GGGCGATATTCGTTAATTCAGCGCAATATTAACGCTCAA..... 1568  
  
331 ValThrArgIleAsnAspGluSerArgThr..... 340  
: ||||| : : : : : |||  
1569 .....GCTAGTGGTGATACGCTAAACCCGGTGGTTTGTGGAGACGTC 1612  
  
341 .....ValAspVal..GluT 345



878 CGAAACATCAAGCGCGTAATTTTCCACCTTCGAGCAAAACCAAGATAAAG 927  
 100 euIleAlaGluIleAsnPheGluGlyAsnArgLeuIleProLysGluGly 116  
 928 CGCTCGGTGAATTTGTAATCAGGTTTAATTACTGTCGTAAGACGCG 977  
 117 .....LeuGlnGluGlyLeuLysAsnAlaGly..... 125  
 978 AGTGTAATCTTATTGGTGGCAAGTGAATAACGAGGGTGTGATTACCGT 1027  
 126 .....LeuAlaValGlyGlnProLeuLysGlnAla 136  
 1028 AAATGGTGGCAGCATTTCTTACTCGCAGGGCAAAATC.....A 1068  
 136 hrValGlnMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGly 152  
 1069 CCATCAGCGATATA..... 1082  
 153 TyrTyrAsnThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArg 169  
 1083 ...ATAAACCCCAACCATTTACTTACAGCATTCGCCCGCTGAAAATGAAGC 1129  
 169 gValLysLeuAspMetThrPheAlaGluGlyLysProAlaArgValVala 186  
 1130 GGTCATCTGGCGATATTTTGGCAAGCGGT.....A 1164  
 186 spIleAsnIle.....IleGlyAsnGlnHisPheSerAspAla 198  
 1165 ACATTAATGTCGGTGGCCATCTTACAGCATTCGCCCGCTGAAAATGAAGC 1214  
 199 AspLeuIleAspValLeuAlaIleLysAspAsnLysIleAsn...ProLe 214  
 1215 GATTCGTAAGC.....AAAGATAAAGCGGCAATATTGTCT 1252  
 214 uSerLysAlaAspArgTyrThrGlnGluLysLeuValThrSerLeuGluA 231  
 1253 TTTCCGCGCAAGAGGTGAAGCGGAATGTCGGGTGTAAATTCGCTCAA 1302  
 231 snLeuArgAlaLysTyrLeuAsnAlaGlyPheValArgPheGluIleLys 247  
 1303 ATCAGCAAGCT.....AAA 1316  
 248 AspAlaLysLeuAsnIleAsnGluAspLysAsnArgIle..... 260  
 1317 GCGCGCAAGCTGATGATTACAGCGCATTAAGCTCATTAATAAACAGGTGC 1366  
 261 .PheValGluIleSerLeuHisGlu...GlyGluGlnTyrArgPheGlyG 276  
 1367 AGTTATCGACCTTTCAGGTAAAGAAGGGGAGAACTTACCTTGGCGGTG 1416  
 276 lnThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAlaGluLeuGluAla 292  
 1417 ACGAGCGCGCGAAGGT.....AAAAAGGGCATTCATTA 1451  
 293 LeuLeuLysPheLysAlaGluGluGlyPheSerGlnAlaMetLeuGluG 309  
 1452 GCAAGAAACCTCTTTAGAAAAAGGC..... 1478  
 309 nThrThrAsnIleSerThrLys.....P 318  
 1479 .TCAACCATCAATGATATCAGCAAGAAAGAAAGCGCGCGCTATTGTGT 1527  
 318 heGlyAsp.....AspGlyTyrTyrTyrAlaGlnIleArgPro 330  
 1528 GGGCGGATATGCGTTAATTCACGCGCAATATTAAACGCTCAA..... 1568  
 331 valThrArgIleAsnAspGluSerArgThr..... 340  
 1569 .....GGTAGTGGTGATATCGCTAAAAACCGGTGTTTGTGGAGACGTC 1612  
 341 .....ValAspVal...Glu 345

```

alignment_scores:
  Quality: 147.00      Length: 939
  Ratio: 0.347         Gaps: 53
  Percent Similarity: 45.154      Percent Identity: 20.554

alignment_block:
  US-09-701-711-2 x US-08-728-470-1  ..

Align seg 1/1 to: US-08-728-470-1 from: 1 to: 5116

27  GlnAlaAlaAspPheMetAlaAsnAspIleThrIleThrGlyLeuGlnAr 43
    :: :::::::::::::::Thr :: :: :::::::::::
681  GAAATGGTGCAGTTTTTACAAAGAAACAAACAACTCCGCCGATTCAACGC 730
    :::::::::::::::

43  gValThr.....IleGluSerLeuGlnSerValLeuPropheArgL 57
    |||||
731  TGTTTACATCTAAACCAATCTCCCAATTTAAAGGGATTTTAGATTCTAAC. 779
    |||||

57  euGlyGlnVal.....ValSerGluAsnGlnLeuAlaAspGlyValLys 71
    |||||
780  ..GGACAAGTCTTTTAAATCAACCAATAGGTATCACAATAGTAGTAAGAC 827
    |||||

72  AlaLeuTyrAlaThrGlyAsnPheSer.....AspValGlnVa 84
    |||||
828  GCAATTTATTAACACTAATGGCTTTTACGGCTTCTACGCTAGACATTTCTAA 877
    |||||

84  ITvrHisGlnGluGlyArg...IleIleTyrGlnValThrGluArgProL 100

```



```

; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/617,697
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 05-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-557
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5116 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-617-697-1

```

```

alignment_scores:
  Quality: 147.00      Length: 939
  Ratio: 0.347         Gaps: 53
  Percent Similarity: 45.154      Percent Identity: 20.554

```

```
alignment_block:
US-09-701-711-2 x US-08-617-697-1 ..

```

```
Align seg 1/1 to: US-08-617-697-1 from: 1 to: 5116

```

```

27 GlnAlaAlaAspPheMetAlaAsnAspIleThrIleThrGlyLeuGlnAr 43
   :::::::::::::::::::::
681 GAATGGTGAGTGTATACAGAAACACAACTCCGCGTATTAACCG 730
   :::::::::::::::::::::
43 gValThr.....IleGluSerLeuGlnSerValLeuProPheArgL 57
   |||||
731 TGTATCATCTAACCAATCTCCCAATTAAGGATTTAGATTCTAAC. 779
   |||||
57 euGlyGlnVal.....ValSerGluAsnGlnLeuAlaAspGlyValLys 71
   |||||
780 ..GGACAAGTCTTTTAAATCAACCAATGGTATCAATAGGTAAAGAC 827
   |||||
72 AlaLeuTyrAlaThrGlyAsnPheSer.....AspValGlnVa 84
   |||||
828 GCAATTATTAACTATGCTTTACGGCTCTACGCTAGACATTCTTAA 877
   |||||
84 lTyrHisGlnGluGlyArg...IleIleTyrGlnValThrGluArgProL 100
   :::::::::::::::::::::
878 CGAAACATCAAGCGCGTAAATTTACCTTCGAGCAAAACCAAGATAAG 927
   :::::::::::::::::::::
100 eutIleAlaGluIleAsnPheGluGlyAsnArgLeuIleProLysGluGly 116

```

```

928 CGCTCGTGAATTCACGGTTTAATTACTGTGCTAAAGACGGC 977
   :::::::::::::::::::::
117 .....LeuGlnGluGlyLeuLysAsnAlaGly..... 125
   :::::::::::::::::::::
978 AGTGTAAATCTATTGTGGCAAAAGTGAACACGAGGCTGTATTAGCGT 1027
   :::::::::::::::::::::
126 .....LeuAlaValGlyGlnProLeuLysGlnAlaT 136
   |||||
1028 AAATGGTGGCAGCATTTCTTTACTCGCAGGGCAAAATC.....A 1068
   |||||
136 hrValGlnMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGly 152
   |||||
1059 CCATCAGCGATATA..... 1082
   |||||
153 TyrTyrAsnThrGluIleThrValLysGlnThrMetLeuAspGlyAsnAr 169
   |||||
1083 ...ATAACCCCAACCATTTACTACAGCATTCGCCGCGCTGAAATGAAGC 1129
   |||||
169 gValLysLeuAspMetThrPheAlaGluGlyLysProAlaArgValValA 186
   |||||
1130 GGTCAATCTGGCGGATATTTTTCGCAAAAGCGGT.....A 1164
   |||||
186 spIleAsnIle.....IleGlyAsnGlnHisPheSerAspAla 198
   :::::::::::::::::::::
1165 ACATTAATGTCCGTGCTGCCACTATTCGAAACCAAGGTAAACTTTCTGCT 1214
   |||||
199 AspLeuIleAspValLeuAlaIleLysAspAsnLysIleAsn...ProLe 214
   |||||
1215 GATTCTGTAAGC.....AAAGATAAAAGCGCAATATTGTCTT 1252
   |||||
214 uSerLysAlaAspArgTyrThrGlnGluLysLeuValThrSerLeuGluA 231
   |||||
1253 TTCCGCCAAAGAGGGTGAAGCGGAATTCGCGGTGTAATTCGCTCAAA 1302
   |||||
231 snLeuArgAlaLysTyrLeuAsnAlaGlyPheValArgPheGluIleLys 247
   |||||
1303 ATCAGCAAGCT.....AAA 1316
   |||||
248 AspAlaLysLeuAsnIleAsnGluAspLysAsnArgIle..... 260
   :::::::::::::::::::::
1317 GCGCGCAAGCTGATGATTACAGCGGATAAAGTCACATTAAAAACAGGTGC 1366
   :::::::::::::::::::::
261 .PheValGluIleSerLeuHisGlu...GlyGluGlnTyrArgPheGlyG 276
   :::::::::::::::::::::
1367 AGTTATCGACCTTCAGGTAAGAGAGGGGAGAACTTACCTTGGCGGTG 1416
   :::::::::::::::::::::
276 lnThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAlaGluLeuGluAla 292
   :::::::::::::::::::::
1417 ACGAGCGCGCGAAGCT.....AAAAAGGCGCATTCATTAA 1451
   :::::::::::::::::::::
293 LeuLeuLysPheLysAlaGluGluGlyPheSerGlnAlaMetLeuGluG 309
   |||||
1452 GCAAGAGAAACCTCTTTAGAAAAAGG..... 1478
   :::::::::::::::::::::
309 nThrThrAsnAsnIleSerThrLys.....P 318
   |||||
1479 .TCACCATCAATGTATCAGCAAGAAAGGCGGCGCTATTGTGT 1527
   :::::::::::::::::::::
318 heGlyAsp.....AspGlyTyrTyrTyrAlaGlnIleArgPro 330
   :::::::::::::::::::::
1528 GGGCGCATATTGCGTTAAATTGACGGCAATATTAAACGCTCAA..... 1568
   |||||
331 ValThrArgIleAsnAspGluSerArgThr..... 340
   :::::::::::::::::::::
1569 .....GGTAGTGTGATATCGCTAAACCGGTGTTTGTGGAGAGCTC 1612
   :::::::::::::::::::::
341 .....ValAspVal...GluT 345
   |||||
1613 GGGCATGATTATTATCATCAAGACAAATGCAATTTGTGACGCCCAAGAGT 1662
   :::::::::::::::::::::
345 yrTyrIleAspPro.....ValHisProValTyrValArgArg 357
   :::::::::::::::::::::

```



```
1663 GGTGTTAGACCCGGGTAATGTATCTATTAAATGCAGAAACAGCAGCAGC 1712
358 IleAsn.....PheThrGlyAsnPhelystThrGlnAs 368
||||:||||:||||:||||:||||:||||:||||:||||:||||:
1713 AGCAATACCTTCAGAACAGCATGAATACAGCGGATCCGGGAATAGTCCAG 1762
368 pGluValLeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerA 385
||||:||||:||||:||||:||||:||||:||||:||||:||||:
1763 CACCCCAAAACGAAACAAA.....GAAAGACAAACATTAAACAA 1800
385 snGlnLysIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePhe 401
|||:||||:||||:||||:||||:||||:||||:||||:||||:
1801 ACACAACCTCTTGAG.....AGTATACATAAAAGGTACCTTT 1838
402 LysHisValThrValAspThrArgProValProAsnSerProAspGlnVa 418
||||:||||:||||:||||:||||:||||:||||:||||:||||:
1839 GTTAACATCAGCTGCTAATCAACGCATCTATGCAATAGC..... 1877
418 lAspValAsnPhelValValGluGlnProSerGlySerSerThrIleA 435
||||:||||:||||:||||:||||:||||:||||:||||:||||:
1878 .TCCATTAATTTA.....TCCAAATGCGCAGCTTAACCTTT 1911
435 laAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSer 451
|||:||||:||||:||||:||||:||||:||||:||||:||||:
1912 GGAGTCAGGGTCGGAGCGGTGGCGGCTTGAGATT..... 1946
452 GlnAsnAsnPhemMetGlyThrGlyLysHisValAsnAlaSerPheSerAr 468
||||:||||:||||:||||:||||:||||:||||:||||:||||:
1947 ..AACACGATATTACCACCGGTGATGAT.....ACCAG 1978
468 gSerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheT 485
|||:||||:||||:||||:||||:||||:||||:||||:||||:
1979 AGGTGCAAACTTAACAATTTACTCAGCGCGCTGGGTGTGATTCATAAAA 2028
485 hrValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLys 501
||||:||||:||||:||||:||||:||||:||||:||||:||||:
2029 ATATCTCAGCTCGGGCGCAAGGTAAACATTAACAGCTAACAAAGAT 2078
502 TyrAspAsnLysAsnIleSerAsnTyrValLeuAspSerTyrGlyGlySe 518
||||:||||:||||:||||:||||:||||:||||:||||:||||:
2079 ATCGCCTTTGAGAAAGGACCAACCAAGTCATTTACAGGTCAA....GGGAC 2125
518 rLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSerPheGlyL 535
||||:||||:||||:||||:||||:||||:||||:||||:||||:
2126 TATTACTCTCAGGC.....AATCAAAAA.....GGTT 2151
535 euAsnAlaAspAsnThrLysLeuHisGly..... 544
2152 TTAGATTTAATAATGTCTCTAAACGGCAGCTGGCAGCGAGCTGCAATTC 2201
545 .....GlyArgPheMetGlyI1 550
||||:||||:||||:||||:||||:||||:||||:||||:||||:
2202 ACCACTAAAGAAACCAATAATACGCTATCACAAATAAATTTCAAGGGAC 2251
550 eSerAsnValLysGlnLeuMetAlaAspGlyGlyLysIleGlnValAspa 567
||||:||||:||||:||||:||||:||||:||||:||||:||||:
2252 TTTAAATATT.....TCAGGAAAGTGAACATC...T 2280
567 snAsnGlyIleProAspPheLysHisAspTyrThrThrTyrAsnAlaIle 583
|||:||||:||||:||||:||||:||||:||||:||||:||||:
2281 CAATGTTTACCCTAAATAAGAAAGGATATGATAAATTCAAAGGACGC 2330
584 LeuGlyTyrAsnTyrSerSerLeuAsp..... 592
||||:||||:||||:||||:||||:||||:||||:||||:||||:
2331 ACTTACTGGAATTTAACCTCCTTAATGTTCGAGAGTGGCGAGTTTAA 2380
593 .....Argp 594
||||:||||:||||:||||:||||:||||:||||:||||:||||:
2381 CCTCACTATTGACTCCAGAGGAAGCATAGTCAGGACACACTTACCAGC 2430
594 roValPheProThrGlnGlyMetSerHisSerValAspLeuThrValGly 610
|||:||||:||||:||||:||||:||||:||||:||||:||||:
2431 CT...TATAATTTAAACGGTATCATTCATCAACAAAGACACTACC..... 2471
```

```
611 PheGlyAspLysThrHisGlnLysValValTyrGlnGlyAsnIleTyrAr 627
||||:||||:||||:||||:||||:||||:||||:||||:||||:
2472 TTTAATGTTGAACGAATTCGAGAGTCAACTTTGAC..... 2507
627 gProPheIleLysLysSerVal...LeuArgGlyTyrAlaLysLeuGlyT 643
||||:||||:||||:||||:||||:||||:||||:||||:||||:
2508 .....ATCAAGGCACCAATAGGATAAATAAGTATCTTAGTTGAATT 2550
643 yrGlyAsnAsnLeuProPheTyrGluAsnPhetTyrAlaGlyGlyTyrGly 659
||||:||||:||||:||||:||||:||||:||||:||||:||||:
2551 ACCGATCA.....TTAATGGAACACATTTTCAGTTTCGGAGGGGG 2591
660 SerValArgLysTyrAspGlnSerSerLeuGlyProArgSerGlnAla... 675
||||:||||:||||:||||:||||:||||:||||:||||:||||:
2592 AGTGTT.....GATTTACACCTTCGCCTCATCTCTAACGTCCA 2632
676 .....TyrLeuThrAlaArgGlyG 683
||||:||||:||||:||||:||||:||||:||||:||||:||||:
2633 AACCCCGGTGTAGTTAATAAATCTAAATACTTTAATGTTTCAACAGGT 2682
683 lnGln.....ThrThrLeuGlyGluValValGlyGlyAsnAlaLeu 696
||||:||||:||||:||||:||||:||||:||||:||||:||||:
2683 CAAGTTTAAGATTTAAACCTTCAGGCTCAACAAAAACTGGCTTCTCAAT 2732
697 .....AlaThrPheGlySerGluLeuLeuPr 706
||||:||||:||||:||||:||||:||||:||||:||||:||||:
2733 GAGAAAGATTTAATTAATGCCACCGGAGGCACATAACACTTTTGCA 2782
706 oLeuProPheLysGlyAspTyrPileAspGln.....ValArgProV 720
|||:||||:||||:||||:||||:||||:||||:||||:||||:
2783 AGTTGAAGCACCAGATGGAATGATTGTTAAAGCATTTGTAGCCAAAAA 2832
720 aIlePheIleGluGlyGlnValPheAspThrThrGlyMetAspLys 736
|||:||||:||||:||||:||||:||||:||||:||||:||||:
2833 ACATAACCTTTGAAGGAGGTAAACATC.....ACCTTTGGCTCCAGGAA 2876
737 GlnThrIleAspLeuThrGlnPheLysAspProGlnAlaThrAlaGlu 753
|||:||||:||||:||||:||||:||||:||||:||||:||||:
2877 GCGGTAAACAGAAATC.....GAAGCAATGTTACTATCAATAA 2914
753 nAsnAlaLysAlaAla.....AsnArgp 761
||||:||||:||||:||||:||||:||||:||||:||||:||||:
2915 CAACGCTAACGCTCCTTATCGTTTCGGATTTTGACACCATCAAAAC 2964
761 roLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAla 777
||||:||||:||||:||||:||||:||||:||||:||||:||||:
2965 CTTTAATATTTAAAGATGTCATCATTAATAGCGGCAACCTTACCGCT 3014
778 ThrTrpTyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLe 794
||||:||||:||||:||||:||||:||||:||||:||||:||||:
3015 .....GGAGCAATATTTGCAATATAGCCGGAATCT 3046
794 uAsnLysLysGlnAsnAspGlnThrAspThrVal.....GlnPheG 808
||||:||||:||||:||||:||||:||||:||||:||||:||||:
3047 TACCGTTGAAGTAACGCTAATTTCAAGCATATCACAAATTTTCACCTTTA 3096
808 lnIleGlySerValPhe 813
||||:||||:||||:||||:||||:||||:||||:||||:||||:
3097 ATGTAGCGCGCTTGTTT 3113
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:us-08-719-641-1
seq_documentation_block:
; Sequence 1, Application US/08719641
; Patent No. 6218141
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
```

Thu Sep 19 10:04:26 2002

```

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/719,641
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-625
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5116 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-719-641-1

```

```

alignment_scores:
Quality: 147.00 Length: 939
Ratio: 0.347 Gaps: 53
Percent Similarity: 45.154 Percent Identity: 20.554

alignment_block:
US-09-701-711-2 x US-08-719-641-1 ..

Align seg 1/1 to: US-08-719-641-1 from: 1 to: 5116

27 GlnAlaAlaAspPheMetAlaAsnAspIleThrIleThrGlyLeuGlnAr 43
::: :::::||||| ::::: ::::: ::::: ::::: :::::
681 GAAATGGTGCAGTTTACAAAGAACACACACATCCGCGCTATTAACCG 730
43 gValThr.....IleGluSerLeuGlnSerValLeuProPheArgL 57
||||| ::::: ::::: ::::: ::::: :::::
731 TGTACATCTAACCAATCCCAATTAAGGATTTAGATCTTAAC. 779
57 euGlyGlnVal.....ValSerGluAsnGlnLeuAlaAspGlyValLys 71
||||| ::::: ::::: ::::: ::::: :::::
780 ..GGACAAGTCTTTTAAATCAACCCCAATGGTATCAATAGGTAAAGAC 827
72 AlaLeuTyrAlaThrGlyAsnPheSer.....AspValGlnVa 84
||||| ::::: ::::: ::::: ::::: :::::
828 GCAATATTAACTAATAGCTTTACGGCTTCTACGTAGACATTTCTAA 877
84 lTyrHisGlnGluGlyArg...IleIleTyrGlnValThrGluArgProL 100
::: ::::: ::::: ::::: :::::
878 CGAAACATCAAGCGCGTAAATTTTCACTTCGACCAACCAACCAATAAG 927
100 euIleAlaGluIleAsnPheGluGlyAsnArgLeuIleProLysGluGly 116
::: ::::: ::::: ::::: :::::

```

```

928 CGTCGCTGAAATTTGAATCAGGTTTAAATTACTGTCGGTAAAGACGCG 977
117 .....LeuGlnGluGlyLeuLysAsnAlaGly..... 125
::: :::::||||| :::::
978 AGTGTAATCTTATTGTTGCAAAAGTGAAACGAGGTTGTGATTAGCGT 1027
126 .....LeuAlaValGlyGlnProLeuLysGlnAla 136
||| :::::||||| :::::
1028 AAATGGTGGCAGCATTTCTTTACTCGCAGGCAAAATC.....A 1068
136 hrValGlnMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGly 152
||:|||||
1069 CCATCAGCATATA..... 1082
153 TyrTyrAsnThrGluIleThrValLysGlnThrMetLeuAspGlyAsnAr 169
||| ||||| :::::
1083 ...ATAACCCCAACCATTTACTTACAGCATTCGCCGCTGAAATGAAGC 1129
169 gValLysLeuAspMetThrPheAlaGluGlyLysProAlaArgValVala 186
||:|||||
1130 GGTCAATCTGGCGATATTTTTCGCAAGGCGGT.....A 1164
186 spIleAsnIle.....IleGlyAsnGlnHisPheSerAspAla 198
||:|||||
1165 ACATTAATGTCGTCGCCACTATTCCAAACCAAGGTAAACTTCTGCT 1214
199 AspLeuIleAspValLeuAlaIleLysAspAsnLysIleAsn...ProLe 214
||| :::::
1215 GATTCTGTGAAGC.....AAGATAAAGCGGCAATATTGTTCT 1252
214 uSerLysAlaAspArgTyrThrGlnGluLysLeuValThrSerLeuGluA 231
||| :::::
1253 TTCCGCCAAAGAGGTGAAGCGGAATTTGGCGGTGTAATTTCCGCTCAA 1302
231 snLeuArgAlaLysTyrLeuAsnAlaGlyPheValArgPheGluLys 247
|| :::::
1303 ATCACAAGCT.....AAA 1316
248 AspAlaLysLeuAsnIleAsnGluAspLysAsnArgIle..... 260
1317 GCGCGCAAGCTGATGATTACAGCGCATTAAGTCACATTAAAAACAGGTGC 1366
261 .PheValGluIleSerLeuHisGlu...GlyGluGlnTyrArgPheGlyG 276
1367 AGTTATCGACCTTTTACGTAAGAGAGGGGAGAACTTACCTTGGCGGTG 1416
276 lntThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAlaGluLeuGluAla 292
::: :::::
1417 ACGAGCGCGCGAAGGT.....AAAAGGCGCATTCATTA 1451
293 LeuLeuLysPheLysAlaGluGlyPheSerGlnAlaMetLeuGluG 309
||| :::::
1452 GCAAGAAAAACCTCTTTAGAAAAAGGC..... 1478
309 nThrThrAsnAsnIleSerThrLys.....P 318
1479 .TCAACCATCAATGTATCAGGCAAAAGAAAGCGGCGCTATTGTGT 1527
318 heGlyAsp.....AspGlyTyrTyrAlaGlnIleArgPro 330
::: |||||
1528 GGGCGCATATTGCGTTAATTGACGCAATATTAAACGCTCAA..... 1568
331 ValThrArgIleAsnAspGluSerArgThr..... 340
1569 .....GGTAGTGGTATATCGCTAAAAACCGTGTGTTTGTGGAGAGCTC 1612
341 .....ValAspVal...GluT 345
1613 GGGCGCATGATTATTATCATCAAGACAAATGCAATGTTGACGCCCAAGAGT 1662
345 yTyrIleAspPro.....ValHisProValTyrValArgArg 357
::: :::::
1663 GGTGTTAGACCCCGGATAATGATCTATTAATGCAGAAACAGCAGGACGC 1712

```



```

alignment_scores:      Length: 704
      Quality: 146.50      Caps: 37
      Ratio: 0.467
Percent Similarity: 44.602 Percent Identity: 21.449

alignment_block:
US-09-701-711-2 x US-08-968-685A-9 ..

Align seg 1/1 to: US-08-968-685A-9 from: 1 to: 9542

34 AsnAspIleThrIleThrGlyLeuGlnArgValThrIleGlu..... 47
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2648 ATGGTATAGCACTTGACGGCACACTTACATCATCCACAAAGACAAGTTGG 2697
   48 .....SerLeuGlnSerValLeuProPhe..... 55

2698 CTTTGCTAAGCAAGATGGTTTCATCTTATATAAGCAAAACCTTATCTTGATA 2747
   56 .....ArgLeuGlyGlnVal...ValSerGluAsnGlnLeuAla 67
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2748 AGGACAAGCTAAAGTGGTGAAGTTGAGATTACCACCAAGCGCATTAAT 2797
   68 AspGlyValLysAlaLeuTyraIleThrGlyAsnPheSerAspValGlnVa 84
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2798 GCAGGTGGTAAAGCCATCATCAGGACTAAGCAATACCTTACCAGCATGCCAC 2847
   84 lTyrHisGlnGluClyArgIle.....IleTyrGlnValThrG 97
   : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2848 CAACGCAACCAACAGCGCATCTACTCAATGGGTATCGTTGATGAGTACTG 2897
   97 luArgProLeuIleAlaGluIleAsnPheGluGlyAsnArgLeuIlePro 113
   : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2898 ACAAACCCGTCGCCGACGATGGT..... 2923

114 LysGluGlyLeuGlnGluGly.....LeuLysAsnAlaGlyLeuAlaVa 128
   : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2924 ...GATGTGCTAAACGCGAGCGCTTAACTCTAAAAAATAATGTTGACGCCAA 2970
   128 lGlyGlnProLeuLysGlnAlaIleThrValGlnMetIleIleThrGluLeu 145
   : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2971 AGACTTTGTCTCCACTTATGACACTGTGTGATTTTATCAATGGCAATGCCA 3020

```

262 lgluileSerLeuHisIuGluGlyGluGlnTyrArgPheGlyGlnThrGlnP 279  
::: |||::: :::: ::|||::: :::  
3367 CACCGTG.....GGTAAAGATGCAAAACCAATCAAGTCAACA 3404  
::: |||::: :::: ::|||::: :::  
279 heLeu..... 280  
|||  
3405 CCCTAAACCTCAAGGTAAACCGGCTTGATATTCAACCAATAAGAT 3454  
281 lGlyAsnLeuThrTyr.....ThrGlnAlaGluLeuGluAla..... 292  
|||::: |||::: :::: ::|||::: :::  
3455 GGTACGGTTACCTTTGGCATTAAACACCAACCGCGTCTTAAAGCCGCCAA 3504  
293 .....LeuLeuLysPheLysAlaGluGluGlyPheSerGlnAlaMetL 307  
|||::: :::: ::|||::: :::  
3505 CAACACCACTCTAAAC.....ACAAATGGCTTGCT.....A 3536  
307 euGluGlnThrThrAsnAsnIleSerThrLysPheGlyAspAspGlyTyr 323  
|||::: |||::: |||::: :::: ::|||::: :::  
3537 TTAANAACACCGCTGGTAAACGAACAATCCAAAGTCGGTCTCATGCGCGTG 3586  
324 TyrTyrAlaGlnIleArg.....Prova 331  
::: |||::: |||::: :::  
3587 AAGTTTGCCAGAGTTAATAATGGTGTGTAGGTGGTGGCATTGATGGCAC 3636  
331 lThrArgIleAsnAspGluSerArgThrValAspValGluTyrTyrIleA 348  
::: |||::: |||::: :::  
3637 AACTCGCAT..... 3646  
348 spProValHisProValTyrValArgArgIleAsnPheThrGlyAsnPhe 364  
3646 ..... 3646  
365 LysThrGlnAspGluValLeuArgArgGluMetArgGlnLeuGluGlyAl 381  
|||::: |||::: |||::: :::  
3647 ..ACCAGAGATGAAATT.....GGCTT 3666  
381 aLeuAlaSerAsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArgT 398  
::: |||::: |||::: |||::: :::  
3667 TGCTGGGACTAATGGCTCACTTGATTAAGCAAAACCCCACTTAAGCAAG 3716

```
398 hrGlyPhe.....PheLysHisValThrValAspThrArgPro 410
|||||
3717 ACGGCATTACCGCAGGTGTGTAAGAAATACCAACATTCATCAGGTGAG 3766
|||||
411 Val...ProAsnSerProAspGlnVal.....AspVa 420
::: ||||| ||| |||
3767 ATGGCCAAACACGCAATGATGCTGTGACAGCGGCGCAAGATTATGATTT 3816
|||||
420 lAsnPheValValGluGluGlnProSerGlySerSerThrIleAlaAlaG 437
::: |||||
3817 AAAACCCAGCACTTGAACAAATCAGCATCTGCCAAACAGCACAAA 3866
|||||
437 lyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSerGlnAsn 453
::: ||||| ||| |||
3867 ACTCATTCACGCAATCTCAGTAGCA.....GATCAACAAGGTAAT 3907
|||||
454 AsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerArgSerG 470
|||||
3908 AACTTT.....ACGGTTAGTAACCTTACTCCAGTTATGACACCTCAA 3951
|||||
470 uThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheThr...V 486
::: |||||
3952 GACCTCTGATGTCATCACCCTTGCAGGTGAAACAGGCATTACCAACG 4001
|||||
486 alAsn.....GlyValSerGlnSerLeuSerGlyTyr 496
|||||
4002 TAAATAAAGGTGTGGTGCCTGTGGGCATTGACCAACCAAGGCTTAACC 4051
|||||
497 TyrArgLysThrLysTyrAspAsnLysAsnIleSerAsnTyrValLeuAs 513
||| |||||
4052 ACGCCTAAGCTGACCGTGGGTAAATAATATGCAAGGCAATGTCATTGA 4101
|||||
513 pSerTyrGlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnA 530
|||||
4102 CAGCCAA..... 4108
|||||
530 rgIleSerPheGlyLeuAsnAlaAspAsnThrLysLeuHisGlyGlyArg 546
|||||
4109 .....AATGTCAAAATACC..... 4123
|||||
547 PheMetGlyIleSerAsn...ValLysGlnLeuMetAlaAspGlyGly 562
::: |||||
4124 ATCAGCAGGACTAAGCAACACTCTAGCTAATGTTACCAATGATAAAGGTAG 4173
|||||
562 stIleGlnValAspAsnAsnGlyIleProAspPheLysHisAspTyrThrT 579
|||||
4174 CGTACGCCACACAGACAGCGCAAGATAATCAACAGCAAGCAAAACCC 4223
|||||
579 hrTyrAsnAlaIleLeu.....GlyTrpAsnTyrSerSer 590
::: |||||
4224 GTGCCGCCAGCATTTGTGATGTGCTAAGCGCAGGCTTTAAC..... 4264
|||||
591 LeuAspArgProValPheProThrGlnGlyMetSerHisSerValAspLe 607
|||||
4265 .....TTGCAAGGCAATGGTGAAGCGGTTGACTT 4293
|||||
607 u.....ThrValGlyPheGlyAsp..... 613
|||||
4294 TGTCTCCACTTATGACACTGTCACTTTGCCGATGGCAATGCCACACCG 4343
|||||
614 .....LysThrHisGlnLysValValTyr 621
|||||
4344 CTAAGGTGACCTATGATGACACACAGCAAAACC...AGTAAAGTGTCTAT 4390
|||||
622 GlnGlyAsnIle 625
::: |||||
4391 GATGTCAAATGTG 4402
```

seq\_name: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:US-09-627-376-3

seq\_documentation\_block:

; Sequence 3, Application US/09627376

; Patent No. 6342385

```
; GENERAL INFORMATION:
; APPLICANT: Qi, Fengxia      Caulfield, Page      Chen, Ping
; TITLE OF INVENTION: MUTACIN I BIOSYNTHESIS GENES AND PROTEINS
; FILE REFERENCE: UAB-17402/22
; CURRENT APPLICATION NUMBER: US/09/627,376
; CURRENT FILING DATE: 2001-05-30
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 15567
; TYPE: DNA
; ORGANISM: Streptococcus mutans
; US-09-627-376-3
```

```
alignment_scores:
  Quality: 145.00      Length: 997
  Ratio: 0.355        Gaps: 45
  Percent Similarity: 40.923      Percent Identity: 17.452
```

alignment\_block:

US-09-701-711-2 x US-09-627-376-3 ..

Align seg 1/1 to: US-09-627-376-3 from: 1 to: 15567

```
53 LeuProPheArgLeuGlyGlnValValSerGluAsnGlnLeuAlaAspG 69
||| |||||
2151 TTGACAGTCATGTTAGCAGAGAGATGACGATTTTCATTTCAAGATT 2200
|||
69 yValLysAlaLeuTyrAlaThrGlyAsnPheSerAsp..... 81
|||
2201 TTTTATGTACAGAAACCACTTACGCACTTTCTAATTTTCTTAGTATA 2250
|||
82 .....ValGlnValTyrHisGlnGluGlyArgIleIle 92
|||
2251 CTGATATGATGGATCCTATTGAATTATTACATAATGAT..... 2288
|||
93 TyrGlnValThrGluArgProLeuIleAlaGluIleAsnPheGluGly 109
|||
2289 .....CCGATATTGTCTGAA..... 2303
|||
109 nArgLeuIleProLysGluGlyLeuGlnGluGlyLeuLysAsnAlaGly 126
|||
2304 .....GGG 2307
|||
126 euAlaValGlyGlnProLeuLysGlnAlaThrValGlnMetIleGluThr 142
::: |||||
2308 TATATTGGCTTCCCATCTCTTAGATCATCTATAATAAATTAGAGAAT 2357
|||
143 GluLeuThrAsn.....G 147
|||||
2358 CAGATTGCAAGTACTAAGGAAAAAAGAATGCAAAAGAGACTATTTTCA 2407
|||
147 nTyrIleSerGlnGlyTyrTyrAsnThrGluIleThrVal..... 160
|||
2408 ATACATGCCCCGT.....TATAACACAGAGATCAACTCCGTTTGGCTGT 2451
|||
161 .....LysGlnThr 163
|||||
2452 TTTCTGCCATCGGAATAGTGGTTTTTCGAACCCCTAGGAAGAGAAA 2501
|||
164 MetLeuAspGlyAsnArgValLysLeuAspMetThrPheAla..... 177
::: |||||
2502 TCTTGTATTGAAAAATCTCTTAATGTGTATCTTTTGGGCTTATAAAGT 2551
|||
178 .....GluGlyLysProAlaArgValValAspIleAsnIle 190
|||
2552 ACGAGATAAACTAGAAAGTATCGCTGAAATTTTAAATACTTTTAAAGTAG 2601
|||
190 leGlyAsn..... 192
|||||
2602 TTGCTAATAATGCTTTGCAAAAGTCAATGATTTTGGCTTTTAGATACA 2651
```









```

; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-557
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4794 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-617-697-7

```

```

alignment_scores:
  Quality: 136.50      Length: 886
  Ratio: 0.324         Gaps: 50
  Percent Similarity: 47.517  Percent Identity: 19.977

```

alignment\_block:

US-09-701-711-2 x US-08-617-697-7 ..

Align seg 1/1 to: US-08-617-697-7 from: 1 to: 4794

```

1 MetArgAsnSerTyr.....PheLysGlyPheGlnValse 12
  ::::::::::::::::::::
274 ATCCGTAATAGCTCAATGCTATCAATTTGGAACAATTTAACATTTGA 323
  ::::::::::::::::::::
12 rAlaMetThrMetAlaValMetMetValMetSerThrHisAlaGlnAla 29
  ::::::::::::::::::::
324 CCAAAATGAATGGTGCAGTTTTTACAAGAAAGCAGCAACTCTGCCGTTT 373
  ::::::::::::::::::::
29 laAspPheMetAlaAsnAspIleThrIleThrGlyLeuGlnArgValThr 45
  ::::::::::::::::::::
374 TCAACCGTGTATCATCTGAC.....CAA 396
  ::::::::::::::::::::
46 IleGluSerLeuGlnSerValLeuProPheArgLeuGlyGlnVal..... 60
  ::::::::::::::::::::
397 ATCTCCCAATTAAGGGATTTAGATTCTAAC.....GGACAAGCTTTT 443
  ::::::::::::::::::::
61 ValSerGluAsnGlnLeuAlaAspGlyValLysAlaLeuTyAlaThrG 77
  ::::::::::::::::::::
444 AATCAACCCAAATGGTATACATAGTAAAGACGCAATTTATTAACTA 493
  ::::::::::::::::::::
77 lyAsnPheSer.....AspValGlnValThrHisGlnGluGly 89
  ::::::::::::::::::::
494 ATGGCTTTACTGCTTACGCTAGACATTTCTAACGAAACATCAAGGCG 543
  ::::::::::::::::::::
90 Arg...IleIleTyrGlnValThrGluArgProLeuIleAlaGluLeAs 105
  ::::::::::::::::::::
544 CGTAATTTACCCCTTGAGCAACACCAAGGATAAGCACTCGCTGAATCGT 593
  ::::::::::::::::::::
105 nPheGluGlyAsnArgLeuIleProLysGluGly.....LeuG 118
  ::::::::::::::::::::
594 GAATCACGCTTTTAAATACCGTTGGTAAAGACGCTAGCGTAAACCTTATTG 643
  ::::::::::::::::::::
118 lnGluGlyLeuLysAsnAlaGly..... 125
  ::::::::::::::::::::
644 GTGGCAAAAGTGAACGAGGCGGTATGCGTAAATGGCGGTAGTATT 693
  ::::::::::::::::::::
126 ...LeuAlaValGlyGlnProLeuLysGlnAlaThrValGlnMetIleG 141
  ::::::::::::::::::::
694 TCTTTACTTGCAGGGCAAAATTC.....ACCATCAGCGATATA... 732
  ::::::::::::::::::::
141 uThrGluLeuThrAsnGlnTyIleSerGlnGlyTyTyTyAsnThrGluI 158
  ::::::::::::::::::::
733 .....ATAAATCCAACCA 745
  ::::::::::::::::::::
158 leThrValLysGlnThrMetLeuAspGlyAsnArgValLysLeuAspMet 174
  ::::::::::::::::::::

```

```

746 TCACTTACAGCATTTGCTGCACCTGAAAAACGAGCATCAATCTGGCGCAT 795
  ::::::::::::::::::::
175 ThrPheAlaGluGlyLysProAlaArgValValAspIleAsnIle..... 189
  ::::::::::::::::::::
796 ATTTTGGCCAAAGGTGGT.....AACATTAAATGTCGCCGC 830
  ::::::::::::::::::::
190 .....IleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspVal 204
  ::::::::::::::::::::
831 TGCCACTATTTCGCAATAAAGGTAACCTTTTCCGCGACTCTGTAAGC.... 876
  ::::::::::::::::::::
204 euAlaIleLysAspAsnLysIleAsn...ProLeuSerLysAlaAspArg 219
  ::::::::::::::::::::
877 .....AAAGATAAAGGTGAACATTCTCTCTGCCAAGAAGGT 918
  ::::::::::::::::::::
220 TyrThrGlnGluLysLeuValThrSerLeuGluAsnLeuArgAlaLysTy 236
  ::::::::::::::::::::
919 GAAGCGGAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCC..... 963
  ::::::::::::::::::::
236 rLeuAsnAlaGlyPheValArgPheGluIleLysAspAlaLysLeuAsnI 253
  ::::::::::::::::::::
964 .....AAAGTGGTAAAGTTGATGA 982
  ::::::::::::::::::::
253 leAsnGluAspLysAsnArgIle.....PheValGluIleSer 265
  ::::::::::::::::::::
983 TTACAGCGCATAAAGTTACATTGAAACGGGTGCAGTTATCGACCTTTTCG 1032
  ::::::::::::::::::::
266 LeuHisGlu...GlyGluGlnTyrArgPheGlyGlnThrGlnPheLeuG 281
  ::::::::::::::::::::
1033 GGTAAAGAGGGGAGAACTTATCTTGGCGGTGACGAGCGTGGCGAAGG 1082
  ::::::::::::::::::::
281 yAsnLeuThrTyrThrGlnAlaGluLeuGluAlaLeuLeuLysPheLys 298
  ::::::::::::::::::::
1083 T.....AAAAACGGCATTCATTAAGCAAAAGAAACCACTT 1117
  ::::::::::::::::::::
298 laGluGluGlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsnIle 314
  ::::::::::::::::::::
1118 TAGAAAAAGGC.....TCAACAATTAATGTG 1143
  ::::::::::::::::::::
315 SerThrLys.....PheGlyAsp..... 320
  ::::::::::::::::::::
1144 TCAGGTAAAGAAAGGTGGCGCGCTATTGTATGGCGCATATTGGGTT 1193
  ::::::::::::::::::::
321 ....AspGlyTyrTyrAlaGlnIleArgProValThrArg..... 333
  ::::::::::::::::::::
1194 AATTGCGCAATATTAAATGCCAAGGTAAGATATCGCTAAAACTGGTG 1243
  ::::::::::::::::::::
334 .....IleAsnAspGluSerArg 339
  ::::::::::::::::::::
1244 GTTTTGGAGACGTGGCGCATTTACTTATCCATTGATGATACGCAATT 1293
  ::::::::::::::::::::
340 ThrValAspValGluTyrTyrIleAspPro.....ValHisPr 352
  ::::::::::::::::::::
1294 GTTAAACAAAGAAAGTGGCTACTAGACCCAGAGAATGTGACTATTGAAGC 1343
  ::::::::::::::::::::
352 oValTyrValArgArgIleAsnPheThrGlyAsnPheLysThrGlnAsp. 368
  ::::::::::::::::::::
1344 TCCTTCGCTCTCGCTCGAGCTGGGTGCGGATAGGAATTCACACTCGG 1393
  ::::::::::::::::::::
369 ..GluValLeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSer 384
  ::::::::::::::::::::
1394 CAGAGGTGATAAAGTACCCCTAAAGAAAAAATAACACCTCTCTGACAA 1443
  ::::::::::::::::::::
385 AsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePh 401
  ::::::::::::::::::::
1444 CTAAACCAATACCACTTTCA.....AATCTCTGAAAGTGGCCACGT 1487
  ::::::::::::::::::::
401 eLysHisValThrValAspThrArgProValProAsnSerProAspGlnV 418
  ::::::::::::::::::::
1488 GGTGAACATAACGCAAGGAGAAACTTACCGTTAATAGCTCTATCAGTA 1537
  ::::::::::::::::::::
418 alaSp.....ValAsnPheValValGluGlnProSerGlySerSer 432
  ::::::::::::::::::::
1538 TAGAAAGAGGCTCCCACTTAATTTCTCCACAGTGAAGGTGAGGGCGGTCAA 1587
  ::::::::::::::::::::

```

```

672 ArgSerGlnAlaTyrLeuThrAlaArgGlyGlnGlnThrThrLeuG1 688
2389 GAAATGATTTAAACTTAACGCC...ACCGTGGCAATATAACAATCAG 2435
688 yCluValValGlyGlyAsnAlaLeuAlaThrPheGlySerGluLeuLe1 705
2436 ACAAGTCGAGGTACCGATTACCGGTCAACAAGAGTGTCCGACCCAAA 2485
705 euProLeuProPheLysGlyAspTrpIleAspGlnValArgProVal1 721
2486 AAAACATAACTTTTAA... 2502
722 PheIleGluGlyGlyGlnValPheAspThrThrGlyMetAspLysGlnTh 738
2503 .....GGGGTAAATATC.....ACCTTCGGCTCTCAAAAAGCCAC 2537
738 rIleAspLeuThrGlnPheLysAspProGlnAlaThrAlaGluGlnAsnA 755
2538 A.....ACAGAAATCAAGGC...AATGTTACCATCAATAAAACA 2575
755 IaLysAla 757
2576 CTAACGCT 2583

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-074-658-69

seq_documentation_block:
; Sequence 69, Application US/09074658
; Patent No. 6184371
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M
; APPLICANT: Run-Pan Du
; APPLICANT: Quijun Wang
; APPLICANT: Yang, Yan-ping
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
; NUMBER OF SEQUENCES: 78
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/074,658
; FILING DATE: 08-MAY-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-795
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 69:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2718 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-074-658-69

alignment_scores:
Quality: 136.00 Length: 882
Ratio: 0.367 Gaps: 45
Percent Similarity: 42.063 Percent Identity: 18.141

```

## alignment\_block:

```

US-09-701-711-2 x US-09-074-658-69
Align seg 1/1 to: US-09-074-658-69 from: 1 to: 2718

9 PheGlnValSerAlaMetThrMetAlaVal.....MetMetVa 21
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
31 TACCAAAACGACCCCTTACCTGGCCATCGCCAGTATTTTGTGCGGT 80
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
21 lMetSerThrHisAlaGlnAlaAlaPheMetAlaAsnAspIleThrI 38
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
81 GGTGATGACAGCGTGGCGCTCTGATGACATCAGCGTCAATGCCACCAATG 130
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
38 lThrGlyLeuGlnArgValThrIleGlnSerLeuGlnSerValLeuPro 54
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
131 TTACCCAACTGCCCAAGGACGGTTTCACCAATACCGAACACAGGTTCAT 180
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
55 PheArgLeuGlnValValSerGlu..... 63
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
181 GACAACACCAATACACCAACCAATCAGGGCAACACACGCGATACAGCAC 230
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
64 .....AsnGlnLeuAlaAspGlyValL 71
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
231 CAGCACAACTGACCCAAATAGCGGATACCACTGACACAAACACAA 280
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
71 ysAlaLeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisGln 87
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
281 AAATGCCCGCGCGAGGGTCTTTGTGATGGGTAAATTCGTGATACC 330
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
88 GluGlyArgIleTyrGlnValThrGluArgProLeuIleAlaGluIl 104
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
331 AGCGAAAAAATGACCCAGATTTATACC...AAAGATTACAGGCGCGT 377
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
104 eAsnPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluGlyL 121
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
378 ACATACAGCAGG.....CAAGGCT 397
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
121 euLysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrVal 137
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
398 TACAGTACTTAGGACCAACAGACGCTGGCCAGATGGCAGGATACAGGT 447
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
138 GlnMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyrTy 154
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
448 AAAAAGCTTACGCCAGCCATCAGACGCTGATGACATTACACCACTTATTT 497
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
154 rAsnThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgValL 171
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
498 TGAT..... 501
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
171 ysLeuAspMetThrPheAlaGluGlyLysProAlaArgValValAspIle 187
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
502 .....AAATCCCCAAAATATCGGATCTG 525
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
188 AsnIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspValLe 204
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
526 CACCTAGAAAAACGAGCATGTGTTGTGATGCTAAAAAAGCAATAACAT 575
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
204 u.....AlaIleLysAspAsnLysIleAsnProLeuS 215
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
576 CAAATATATATGTTATGGTGCATTTGTCATCACCTGCCAAAAGCCACCT 625
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
215 erLysAlaAspArgTyrThrGlnGluLysLeuValThrSerLeuGluAsn 231
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
626 ACATG.....AATTATCAACAAGAA.....CAAAC 651
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
232 LeuArgAlaLys.....TyrLeuAsnAlaGlyPheValAr 243
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
652 ATCAAAAACAAAAACAGCGCGATGATTATCAAAACATTCGTTTGGCTA 701
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
243 gPheGluIleLysAspAlaLysLeuAsn.....IleAsnG 255
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
702 TATGGAGCTAAGAGCTGGACCTAAATAAAAAAAGGTGCAGACACCCAGA 751
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||

```

```

255 luAspLysAsnArg.....IlePheValGluIleSerLeu...HisGlu 268
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
752 GCGACAAGACCGTGCCATCATCTTTACCACACCTACTTTATTTATCAT 801
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
269 GlyGluGlnTyr.....ArgPheGlyGlnThrGlnPheLe 280
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
802 GGTGAGATGCCAGCACCCATCTGCCAAGGCGGTAAATTTGACTATGA 851
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
280 uGlyAsnLeuThrTyrThrGlnAlaGluLeuGluAlaLeuLeuLysPheL 297
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
852 GGGCAATTTGGTTGTAT.....CTGACCGATGTCA 880
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
297 ysAlaGluGluGlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsn 313
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
881 AA.....AAAGCCCATTTTAGATAAAACACAGACGATAAA 915
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
314 lLeSerThrLysPhe.....GlyAsp..... 320
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
916 GTAGGCATTTATTTAACTCAACCAAGAAATCAAAATGAAGCGGATTTGGT 965
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
321 .....AspGlyTyrTyrTyrAlaGlnIleArgP 330
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
966 GAGTGCAGCACACATTTATCTAAACAGCTTTAAATATAAAACACACCCCGG 1015
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
330 roValThrArgIleAsnAspGluSerArgThrVal.....AspValGlu 344
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1016 CCACCTATAGCGTGGACTTTGATCAAAATACCTTAAAGGCAAAATTTGCT 1065
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
345 TyrTyrIleAspProValHisProVal.....TyrValArgAr 357
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1066 TATTATGACAAACCAACAAAGCAACGCGATGGCGTTATATCAGA... 1113
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
357 gIleAsnPheThrGlyAsnPheLysThrGlnAspGluValLeuArgArg 374
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1114 .....AGTCAGTTTGATACCGACAAAAGAGGTCAATGAAGCGG 1150
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
374 luMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnLeu 390
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1151 ATGTCTATGAGATTGACGCCCAAGATTAAATGGCAACCGCTTTACTGGCACA 1200
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
391 SerArgAlaArgLeu.....MetArgThrGlyPhePheLysHisVa 404
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1201 GCCAAATCTTTGATGTGATGATAACACCAATACCGCACTTTTGTGTTAAAGA 1250
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
404 lThrValAspThrArgProValProAsnSerProAspGlnValAspVala 421
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1251 GCTGTTCTCCAAAAGGCCAATCCCAACACCCAGAC..... 1287
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
421 snPheValValGluGlnProSerGlySerSerThrIleAlaAlaGly 437
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1288 .....CCCAACTCAGATACGTAGTAGAGGCGGG 1314
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
438 ...TyrSerGlnSerGlyValThrPheGlnPheAspValSerGlnAs 453
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1315 TTTTATGTTGTCGCGGC.....GATGAGCTGGCGGG 1346
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
453 nAsnPheMetGlyThrGlyLysHisValAsnAlaSerPhe..... 466
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1347 TAAATTTTATCCAATGAC.....AACGCAACTTTTGTGTCCTTTG 1387
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
467 .....SerArgSerGluThrArgGluValTyrSerLeuGlyMetThrAsn 481
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1388 GTGCAACAGGACAAAACG.....ACCGAA 1413
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
482 ProTyrPheThrValAsnGlyValSerGlnSerLeuSerGlyTyrTyrAr 498
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1414 CTTGTCGCCACA..... 1425
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
498 gLysThrLysTyrAspAsnLysAsnIleSerAsnTyrValLeuAspSerT 515
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1426 .AAAACGCTGTAT..... 1437
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||

```

```
515 yrGlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIle 531
1438 .....TTT 1440
532 SerPheGlyLeuAsnAlaAspAsnThrLysLeuHisGlyGlyArgPhe 548
1441 AGTACAGAGATTGAAACCCAGCAGCTGTTGGTGGCAATGAAGAT 1490
548 tGlyIleSerAsnValLysGlnLeuMetAlaAspGlyGlyLysIle... 563
1491 TGGT .....AGCATTATTGACGGTAAAGATTAAATG 1522
564 .....GlnValAspAsnAsnGlyIleProAspPheLysHis 575
1523 ATCAAGTCAATAATCAATGAAGATGAACCTGCCCTGCAGTAATAAA 1572
576 AspTyrThrThrTyrAsnAlaIleLeuGlyTyrPheAsnTyrSerLeuAs 592
1573 GAATATTATGATAT .....AATTATGGACGA ..... 1599
592 pArgProValPheProThrGlnGlyMetSerHisSerValAspLeuThrV 609
1600 .....CCCAACAACAATACCAAAAAATAAAGCCAGCG 1636
609 aIGly .....PheGlyAspLysThrHisGlnLysValVal 620
1637 TCACAAAAAACCCGTATTATTTGGTCAG .....CATGATAGTTTAT 1680
621 TyrGlnGlyAsnIleTyrArgProPheIleLys ..... 631
1681 TTTAATGGTAACTATTATGACTTATCAGCCAAAGAACCAACAAAGCTGG 1730
632 .....LysSerValLeuArgGlyTyr ..... 638
1731 TGTCCTCCCAAGATACCAGCACCATAAGAGTATTTGGCTAAATACCCAG 1780
639 ..AlaLysLeuGlyTyrGlyAsnAsnLeu ..... 647
1781 ATGCCAAAGTAAACAGACAGCAATAAAGTTACCAAAATCGTTCTACAACA 1830
648 .....ProPheTyrGluAsnAsnPheTyrAlaGlyGlyTyrGly 660
1831 GCCAAAGATGAAGCG...TATACCGCATTCATGCC ..... 1863
660 rValArgGlyTyrAspGlnSerSerLeuGlyProArgSerGlnAlaTyrL 677
1864 .....AAAGCTATGACCATCATCAGTTTGGT .....GAAGTATTGT 1900
677 euThrAlaArgArgGlyGlnGlnThrThrLeuGlyGluValValGlyGly 693
1901 ATATGATATAACAAAGGCAACCCCAACAGCAGTATTATTGTGCAAGCGGT 1950
694 AsnAlaLeuAlaThrPheGlySerGluLeuIleLeuProLeuPro ..... 708
1951 CAAGCG .....GATGTCAGTACTCAGCTGCCAGTCG 1982
709 .....PheLysGlyAspTrpIleAspGlnValArgProValI 721
1983 AGGTAATTCACCTATAATGCTCTTTGGCAGGCTAC ..... 2019
721 lePheIleGluGlyGlyGlnValPheAspThrThrGlyMetAspLysGln 737
2019 ..... 2019
738 ThrIleAspLeuThrGlnPheLysAspProGlnAlaThrAlaGluGlnAs 754
2020 .....CTGACCCAGAAAAAGACAAAGAGTTATAGCAAAAGATGAGGA 2060
754 nAla .....LysAlaAlaAsnArgProLeuLeuThrGlnAsp 766
2061 TACCATCAAGCAAAAGGCTCTTAAGATATATATTGACCAAGAC 2106
seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq.us-08-841-178-25
```

```
seq_documentation_block:
; Sequence 25, Application US/08841178
; Patent No. 5880275
; GENERAL INFORMATION:
; APPLICANT: Fischhoff, David A.
; TITLE OF INVENTION: Synthetic Plant Genes and Method for Preparation
; FILE REFERENCE: 38-21(15119)A
; CURRENT APPLICATION NUMBER: US/08/841.178
; CURRENT FILING DATE: 1997-04-29
; EARLIER APPLICATION NUMBER: US 08/433,111
; EARLIER FILING DATE: 1995-05-03
; EARLIER APPLICATION NUMBER: US 07/959,506
; EARLIER FILING DATE: 1992-10-09
; EARLIER APPLICATION NUMBER: US 07/476,661
; EARLIER FILING DATE: 1990-02-12
; EARLIER APPLICATION NUMBER: US 07/315,355
; EARLIER FILING DATE: 1989-02-24
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.0 - beta
; SEQ ID NO 25
; LENGTH: 3534
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:synthetic
; OTHER INFORMATION: structural gene encoding insecticidal protein
; OTHER INFORMATION: derived from B.t.k. HD-73
US-08-841-178-25

alignment_scores:
Quality: 134.50 Length: 716
Ratio: 0.414 Gaps: 35
Percent Similarity: 45.391 Percent Identity: 19.832

alignment_block:
US-09-701-711-2 x US-08-841-178-25 ..
Align seg 1/1 to: US-08-841-178-25 from: 1 to: 3534

5 TyrPheLysGlyPheGlnValSerAla ..... 13
943 TACTGTGTGGACACACAGATCATGGCTCTCCAGTTGGATTGACGGGCC 992
14 ....MetThrMetAlaValMetMetValMetSerThrHisAlaGlnAla 29
993 CGAGTTTACCTTCTCTCTCTGTAACACTATGGAAACGCGCGTCCACAC 1042
29 laAspPheMetAlaAsnAspIleThrIleThrGlyLeuGlnArgValThr 45
1043 AAGTATCGTTGCTCAA .....CTAGTTCAGGGTGTC 1074
46 IleGluSerLeuGlnSerValLeu .....ProPheArgLeuGlyG1 59
1075 TACAGAACCTTGTCTTCCAGCTTGTACAGAACGACCTTCAATATCGGT.. 1122
59 nValValSerGluAsnGlnLeuAla .....AspGlyValLysAlaLeu 74
1123 ....ATCAACAACCAACCAACTTTCCTGTTCTGACGGAACAGAGTCGCCT 1168
74 yrAlaThrGlyAsnPheSerAspValGlnValTyrHisGlnGluGlyArg 90
1169 ATGGAACCTCTCTTAACCTTGCATCGCTGTTTACAGAAAGAGCGGAACC 1218
91 IleIleTyrGlnValThrGluArgProLeuIleAlaGluIleAsnPheG1 107
1219 GTT .....GATTCTTGGACGAATCCACACCACA 1247
107 uGlyAsnArgLeuIleProLysGlyGlyLeuGlnGluGlyLeuLysAsn 124
1248 GAACAACAATGTGCCACCCAGCAAGGATTCTCCACACAGGTTGAGCCACG 1297
```

```

2012  AGTGCAGGCGAGCGAGCTCAGTGATGACGC...AATTACTACCA 2058
398  ThrGlyPhePheLysHisValThrValAspThrArgProValProAsnSe 414
2059  GATTCAAATTCANA..... 2073
414  rProAspGlnValAspValAsnPheValValGluGluGlnProSerGlyS 431
2074  ....GACATTAAT.....AGCAACCCAGAA... 2094
431  erSerThrIleAlaAlaGlyTyrSerGlnSerGlyGlyValThrPheGln 447
2095  ....CGTGGGTGGGGCGAAGTACAGGGATTACCATCCAA 2130
448  ....PheAspValSerGlnAsnAsnPheMetGlyThrGlyLysHisVa 462
2131  GGAGGGGATGACGTATTTAAAGAAAT..... 2157
462  IAsnAlaSerPheSerArgSerGluThrArgGluValTyrSerLeuGlyM 479
2157  .... 2157
479  eThrAsnProTyrPheThrValAsnGlyValSerGlnSerLeuSerGly 495
2158  ....TACGTACACATATCAGGTACCTTTGTAGTGATGCTATCCA 2196
496  TyrTyrArgLysThrLysTyrAspAsn.....LysAsnIleSerAs 509
2197  ACATATTGTATCAAAAAATCGATGAATCAAAATTTAAAGCGCTTTACCG 2246
509  nTyrValLeuAspSerTyrGlyGlySerLeuSerTyrGlyTyrProIleA 526
2247  TTATCAATTAAGAGGTTAT.....ATCG 2269
526  spGluAsnGlnArgIleSerPheGlyLeuAsnAlaAspAsnThrLysLeu 542
2270  AAGATAGTCAAGACTTAGAAATCATTTAAATTCGCTACAATGCAAAACAT 2319
543  HisGlyGlyArgPheMetGlyIleSerAsnValLysGlnLeuMetAlaAs 559
2320  GAAACAGTAATATGCCAGGTACCGGTTCTTATGCGCGCTTTCAGGCCCA 2369
559  poly.....GlyLysIleGlnValAspAsnAsnGlyIleProAspPheL 574
2370  AAGTCCCAATCGGAAAGTGTGGAGAGCCGAATCGATGCGCGCCACAC... 2415
574  yHisAspTyrThrThrTyrAsnAlaIleLeuGlyTyrTpAsnTyrSerSer 590
2416  ....CTGTAATGGAAT...CTGTAC 2433
591  LeuAspArgProValPheProThrGlnGlyMetSerHis..... 603
2434  TTAGATTCTCGTAGGGATGGAGAAAGTGCCTCATCATTCGCATCA 2483
604  ....SerValAspLeuThrValGlyPheGlyAspLysThrHisGlnLysV 619
2484  TTTCTCCTTAGACATGTAGTATGATGATACAGACTTAATAGGAGCACCTAG 2533
619  aIValTyrGlnGlyAsnIleTyrArgProPheIleLysLysSerValLeu 635
2534  GTGATAGG.....GTGATCTTTAGATTAAAGACGCAA 2565
636  ArgGlyTyrAlaLysLeuGlyTyrGlyAsnAsnLeuProPheTyrGlu 651
2566  GATGGCAGCAAGACTAGG.....AATCTAGAGTTTCTCGAA 2604

```

```
seq_documentation_block:
; Sequence 3, Application US/08377690
; Patent No. 5628995
; GENERAL INFORMATION:
```

```

; APPLICANT: PEFFEROEN, Marix
; APPLICANT: JANSSENS, Stefan
; APPLICANT: DENOLF, Peter
; TITLE OF INVENTION: CONTROL OF OSTRINIA
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: The George Mason Bldg., Washington & Prince
; STREET: Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/377,690
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/164,781
; FILING DATE:
; APPLICATION NUMBER: US 07/938,362
; FILING DATE: 31-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Crane-Feury, Sharon E
; REGISTRATION NUMBER: 36,113
; REFERENCE/DOCKET NUMBER: 010830-039
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3537 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Bacillus thuringiensis
; STRAIN: kurstaki HD-73
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3537
; OTHER INFORMATION: /note= "FEATURES: sequence encodes CryIAC
; OTHER INFORMATION: insecticidal crystal protein
; OTHER INFORMATION: PROPERTIES: CryIAC is toxic to Ostrinia nubilalis (among othe
US-08-377-690-3

```

```

alignment_scores:
  Quality: 134.50      Length: 716
  Ratio: 0.414        Gaps: 35
  Percent similarity: 45.391  Percent identity: 19.832

```

```

alignment_block:
US-09-701-711-2 x US-08-377-690-3 ..
Align seg 1/1 to: US-08-377-690-3 from: 1 to: 3537
5 TyrPhelysGlyPheGlnValSerAla..... 13
|||||:|||||:|||||:|||||:
943 TATTGGTCAGGGCATCAATAATGCTCTCCTGTAGGTTTTCGGGGCC 992
14 ....MetThrMetAlaValMetMetValMetSerThrHisAlaGlnAlaA 29
|||||:|||||:|||||:|||||:
993 AGAATTCACATTTCCGCTATATGGAACATATGGAAATCCAGTCCACAAC 1042
29 laAspPheMetAlaAsnAspIleThrGlyLeuGlnArgValThr 45
|||||:|||||:|||||:|||||:

```

```

1043 AACGTATTGTTGCTCAA.....CTAGGTCAGGCGGTG 1074
46 IleGluSerLeuGlnSerValLeu.....ProPheArgLeuGlyG 59
|||||:|||||:|||||:|||||:
1075 TATAGAACATTATCGTCCACTTATATAGAAGACCTTTTATATAGGG.. 1122
59 nValValSerGluAsnGlnLeuAla.....AspGlyValLysAlaLeuT 74
|||||:|||||:|||||:|||||:
1123 ....ATAAATAATCAACAACATCTCTTCTTGACGGGACAGAAATTTGCTT 1168
74 yrAlaThrGlyAsnPheSerAspValGlnValTyrHisGlnGluGlyArg 90
|||||:|||||:|||||:|||||:
1169 ATGGAACCTCCCTCAAAATTTGCCATCGCTGTATACAGAAAGACGGGAGC 1218
91 IleIleTyrGlnValThrGluArgProLeuIleAlaGluIleAsnPheG 107
|||||:|||||:|||||:|||||:
1219 GTA.....GATTGCTGGGATGAATAATACCGCCACA 1247
107 uGlyAsnArgLeuIleProLysGluGlyLeuGlnGluGlyLeuLysAsnA 124
|||||:|||||:|||||:|||||:
1248 GAATAACAACGTCGCCACCTAGGCAAGGATTTAGTCATCGATTAAAGCATG 1297
124 laGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrValGlnMetIle 140
|||||:|||||:|||||:|||||:
1298 TTTCATG...TTTCGTTTTCAGGCTTTAGTAATAGTAGTGTAACTATAA 1344
141 GluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyrTyrAsnThrG 157
|||||:|||||:|||||:|||||:
1345 AGAGCTCTATGTTCTCTTGGATACATCGTAGTGCTGAATTTAAT...AA 1391
157 uIleThrValLysGlnThrMetLeuAspGlyAsnArgValLysLeuAspM 174
|||||:|||||:|||||:|||||:
1392 TATAATTGCATCGGATAGTATTACTCAAAATCCCTGCGTAGGAGGAAACT 1441
174 etThrPheAlaGluGlyLysProAlaArgValValAspIleAsnIleIle 190
|||||:|||||:|||||:|||||:
1442 TTCTTTT.....AATGCTTCTGTAATT 1464
191 GlyAsnGlnHisPheSerAspAlaAspLeuIleAspValLeuAlaIleLy 207
|||||:|||||:|||||:|||||:
1465 TCAGGACCAAGGATTTACTGTGGGGACTTAGTTAGATTAAATAGTAGTGG 1514
207 sAspAsnLysIleAsn.....ProLeu.....S 215
|||||:|||||:|||||:|||||:
1515 AAATAACATTTCAGAAATAGAGGGTATATTGAAGTTCCAATTCACCTCCCAT 1564
215 erLysAlaAspArgTyrThrGlnGluLysLeuValThrSerLeuGluAsn 231
|||||:|||||:|||||:|||||:
1565 CGACATCTACCAGATAT.....CGA 1584
232 LeuArgAlaLysTyrLeuAsnAlaGlyPheValArgPheGluIleLysAs 248
|||||:|||||:|||||:|||||:
1585 GTTCGTGTACGTATGCTTCT.....GTAACCCC 1613
248 pAlaLysLeuAsnIleAsnGluAspLysAsnArgIlePhe..... 261
|||||:|||||:|||||:|||||:
1614 GATTCACCTCAAGCTTAATTTGGGGTAATTCATCCATTTTCCCAATACAG 1663
262 .....ValGluIleSerLeuHisGluGlyGlnTyrArgPheGly 275
|||||:|||||:|||||:|||||:
1664 TACCAGCTACAGCTACGTCATCTAGATAATCTACAATCAAGTGATTGTGGT 1713
276 Gln.....ThrGlnPheLeuGlyAsnLeuThrTy 285
|||||:|||||:|||||:|||||:
1714 TATTTGAAAGTGCCAAATGCTTTTACATCTTCATTAGGTAAATATAGTAGG 1763
285 rThrGln..... 287
1764 TGTTAGAAATTTAGTGGGACTGCAGGAGTGATTAATAGACAGATTTCGAT 1813
288 .....AlaGluLeuGluAlaLeuLeuLysPheLysAlaGlu 299
|||||:|||||:|||||:|||||:
1814 TTATTCAGTTACTGCAACACTCGAGGCTCAATATATATCTGAAAGAGCG 1863

```



3018 AATGATATATTAATACAGGCTTT.....AACCTAAAAATAATAA 3058  
50 nserValLeuProPheArgLeuGlyGlnValValSerGluAsnGlnLeuA 67  
|||||  
3059 CAACCCATTGACTTT.....GTCTCCACTTATGACATTG 3093  
67 laaspGlyValLysAlaLeuTyrAlaThrGlyAsnPheSerAspValGln 83  
|||||  
3094 TTGAC.....TTTGCAATGGCAATGCCACCCGCCACA 3128  
84 ValTyrHis.....GlnGluGlyArgIleIleTyrGlnValTh 96  
|||||  
3129 GTACCCATGATACCGTAACAACACAGTAAGTGGTATATGATGAA 3178  
96 rGluArgProLeuIleAlaGluIleAsnPheGluGlyAsnArgLeuIleP 113  
3179 T..... 3179  
113 roLysGluGlyLeuGlnGluGlyLeuLysAsnAlaGlyLeuAlaValGly 129  
3179 ..... 3179  
130 GlnProLeuLysGlnAlaThrValGlnMetIleGluThrGluLeuThrAs 146  
|||  
3180 .....GTGGATGATACACCACTTCTATCAACAGGCACTGAT..... 3215  
146 nGlnTyrIleSerGlnGlyTyrTyrAsnThrGluIleThrValLysGlnT 163  
3216 .....GACAATAAAAAAACTTGGCGTCAAAACCA 3243  
163 hrMetLeuAspGlyAsnArgValValLysLeuAspMetThrPheAlaGluGly 179  
|||  
3244 CC.....AACTCAACAACAAGTGTATGT 3272  
180 LysProAlaArgValValAspIleAsnIleGlyAsnGlnHisPheSe 196  
|||  
3273 AATACAGCAACTAACTTTAATGTTAAC.....TCTAG 3304  
196 raspAlaasp.....LeuIleaspValLeuAlaIleLysAspAsnLysIleA 212  
|||||  
3305 TGATGAAGATGCCCTTGTTAACGCCAAGACATCGCGGAAT.....CTAA 3351  
212 snProLeuSerLysAlaAspArgTyrThrGlnGluLysLeuValThrSer 228  
|||  
3352 ACACCTAGCAAGGAATTCACACCAACCAAGGACAGCAGACACCGCC 3401  
229 LeuGluAsnLeuArgAlaLysTyrLeu.....As 238  
|||||  
3402 CTACAAACCTTTACCGTTAAAAAGGTAGTGAATAATAATCCTGATGA 3451  
238 nAlaGlyPheValArgPheGluIleLysAspAlaLys..... 250  
|||||  
3452 CGCAACCCATCACCGTGGTCAAAAGACGCAAAATAATCAAGTCAACA 3501  
251 .....LeuAsnIleAsnGluAspLysAsn 258  
|||||  
3502 CCCTAACACTCAAAAGGTGAACCGTCTTAATATTAACACCGCAAAAT 3551  
259 ArgIlePheValGluIleSerLeuHisGluGlyGlnTyrArgPheG 275  
3551 ..... 3551  
275 yGlnThrGlnPheLeuGlyAsnLeuThrTyr.....ThrGlnAlaG 289  
|||||  
3552 .....GGTAGGTTACCTTTGGCATTAAACACCAACAAAGCG 3585  
289 luLeuGluAlaLeuLeuLysPheLysAlaGluGluGlyPheSerGlnAla 305  
|||||  
3586 GTCTTAAGCCGCAAAAGCACCTTAACGACGGTGGCTTGTCT..... 3629  
306 MetLeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAspG 322  
|||||  
3630 ...ATTAACCCACTGGTAGCGAACAAATCAAGTCGGTGTGATGG 3676

322 yTyrTyrTyrAlaGlnIleArg..... 329  
3677 CGTGAAGTTGCCAAGGTTAATAATAATGGTTGGTAGTGCTGCCATTG 3736  
330 ..ProValThrArgIleAsnAspGluSerArgThrValAspValGluTyr 345  
3727 ATGGCACAACCTCGCAT..... 3743  
346 TyrIleAspProValHisProValTyrValArgArgIleAsnPheThrG 362  
3743 ..... 3743  
362 yAsnPheLysThrGlnAspGluValLeuArgArgGluMetArgGlnLeuG 379  
3744 .....ACCAGAGATGAAT..... 3758  
379 luGlyAlaLeuAlaSerAsnGlnLysIleGlnLeuSerArgAlaArgLeu 395  
3759 ..GGCTTTACTGGGACTAATGGCTCCTCTGATAAAAGCAACCCACCTA 3806  
396 MetArgThrGlyPhe.....PheLysHisValThrValAspTh 408  
3807 ACCAAGACGGCATTAACGACGGTGGTAAAAAGATTACCAACATTCAATC 3856  
408 rArgProVal.....ProAsnSerProAspGlnVal..... 418  
3857 AGGTGAGATTGCCAAACAGCCATGCTGTGACAGCGCGCAAGATT 3906  
419 ..AspValAsnPheValValGluGlnProSerGlySerSerThrIle 434  
3907 ATGATTTAAACCACTTGAACAAATCAGCAGTAGTACTGCCAAACA 3956  
435 AlaAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSe 451  
3957 GCACAAACTCATTTACACGAATTCAGTAGCA.....GATGAACA 3997  
451 rGlnAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerA 468  
3998 AGGTAAATAACTTT.....ACGTTAGTAACCTTACTCAGTTATGACA 4041  
468 rGserGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPhe 484  
4042 CCTCAAGACCTCTGATGTCATCACCTTTCAGGTGAAACACGCATTACC 4091  
485 Thr...Valasn.....GlyValSerGlnSerLeuSe 494  
4092 ACCAAGGTAAATAAGGTGGTGGTGGCGGTGTAATAATGCGCAAGGCATTG 4141  
494 rGlyTyrTyrArgLysThrLysTyrAspAsnLysAsnIleSerAsnTyrV 511  
4142 CTTAACCCACCTTAAGCTGACCGTGGTAAATAATGCGCAAGGCATTG 4191  
511 alLeuAspSerTyrGlySerLeuSerTyrGlyTyrProIleAspGlu 527  
4192 TCATTGACAGCCAA.....AATGGTCAAAATACC..... 4205  
528 AsnGlnArgIleSerPheGlyLeuAsnAlaAspAsnThrLysLeuHisG 544  
4206 .....AATGGTCAAAATACC..... 4220  
544 yGlyArgPheMetGlyIleSerAsn...ValLysGlnLeuMetAlaaspG 560  
4221 .....ATCAGAGGACTAAGCAACACTCTAGCTAATGTTACCAATGATA 4263  
560 lyGlyLysIleGlnValAspAsnGlyIleProAspPheLysHisasp 576  
4264 AAGTAGCTGACGACCAACAGAGACAGGCAATATAATCAAGACGACGAC 4313  
577 TyrThrThrTyrAsnAlaIleLeu.....GlyTyrAsnTy 588  
4314 AAAACCCGTGCCCGCAGCATTTGTTGATGCTGAAGCGCAGGCTTTAAC... 4361





[illegible]

4805 AGAAGTTGCCAAAGCAAACTGGTCCCAAGCCCAA..... 4841  
772 erAlaGlyValGlyAlaThrTyrPThrProIleGlyProLeuSer.... 787  
4842 .....ACCCAGATGGCACATTGGCTCAA 4865  
788 IleSerTyrAlaLysProLeuAsnLysLysGln...AsnAsp 800  
.....  
4866 ATGAATGTCAATCAGTCATTAAACAAGAACAAAGTAATGAT 4907

seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-07-671-817A-1

seq\_documentation\_block:  
; Sequence 1, Application US/07671817A  
; Patent No. 5424409  
; GENERAL INFORMATION:  
; APPLICANT: Ely, Susan  
; APPLICANT: Tippet, Janet M  
; TITLE OF INVENTION: DNA constructs  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cushman, Darby and Cushman  
; STREET: Eleventh floor, 1615 L Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036-3601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/671.817A  
; FILING DATE: 19910401  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 8823068.5  
; FILING DATE: 30-SEP-1988  
; PRIOR APPLICATION DATA: PCT/GB89/01157  
; FILING DATE: 29-SEP-1989  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 861-3000  
; TELEFAX: (202) 822-0944  
; TELEX: 6714627 CUSH  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2990 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-07-671-817A-1

alignment\_scores:  
Quality: 130.00 Length: 720  
Ratio: 0.406 Gaps: 35  
Percent Similarity: 44.444 Percent Identity: 19.583

alignment\_block:

US-09-701-711-2 x US-07-671-817A-1 ..  
Align seg 1/1 to: US-07-671-817A-1 from: 1 to: 2990  
5 TyrPheLysGlyPheGlnValSerAlaMetThrMetAla..... 17  
|||||.....  
1023 TATTGGTCAGGGCATCAATATATGCTCTCTGTCGGGTTTCGGGGCC 1072  
18 .....ValMetMetValMetSerThrHisAlaGlnAlaAspPheM 32  
|||||.....  
1073 AGAATTCAGTTTCGGCTATATATGGAACCATGGAAATGCAGTCCA.... 1118

32 etAlaAsnAspIleThrIleThrGlyLeuGlnArgValThrIleGluSer 48  
|||||.....CAACAAGTATTGTGTCTCAACTA 1142  
49 LeuGlnSerValLeu.....Proph 55  
|||||.....  
1143 GGTGAGGGGTGTATAGAACATTATCTCTACTCTTTTATAGAAGACCTTT 1192  
55 eArgLeuGlyGlnValValSerGluAsnGlnLeuAla.....AspGlyV 70  
|||||.....  
1193 TAATATAGG.....ATAAATAACCAACTATCTGTCTTGACGGGA 1236  
70 alLysAlaLeuTyrAlaThrGlyAsnPhSerAspValGlnValTyrHis 86  
|||||.....  
1237 CAGATTTCCTTATGGAACCTCTCAAAATTTGCCATCCGCTGTATACAGA 1286  
87 GlnGluGlyArgIleIleTyrGlnValThrGluArgProLeuIleAlaG 103  
|||||.....  
1287 AAAAGCGGAACGGTA.....GATTCCGCTGGATGA 1315  
103 uLLeAsnPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluG 120  
|||||.....  
1316 AATACCACACAGATAACACAGTGCCTAGGCAAGGATTAGTCATC 1365  
120 lLysLysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThr 136  
|||||.....  
1366 GATTAAAGCCAT.....GTTTCAATGTTTCGTTACGGCTCTAGTAGTAGT 1409  
137 ValGlnMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTy 153  
|||||.....  
1410 GTAAGTATAATAAGAGCTCCTATGTTCTTGGATACATCGTAGTGTGA 1459  
153 rTyrAsnThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgV 170  
|||||.....  
1460 ATTTAAT...AATATAATTGTCATCGGATAGTATTACTCAAAATCCCTGCAG 1506  
170 alLysLeuAspMetThrPheAlaGluGlyLysProAlaArgValValasp 186  
|||||.....  
1507 TGAAGGGAACCTTCTTTT.....AAT 1529  
187 IleAsnIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspVa 203  
|||||.....  
1530 GGTTCGTGTAATTCAGGACCAGGATTACTGTTGGGGACTTAGTTAGATT 1579  
203 lLeuAlaIleLysAspAsnLysIleasn.....ProL 214  
: .....  
1580 AAATAGTAGTGGAAATAACATTCAGAATAAGGGGTATATTGAAGTTCCAA 1629  
214 eu.....SerLysAlaAspArgTyrThrGlnGluLysLeuValThr 227  
: .....  
1630 TTCACCTTCCCATCGACATCTACCAGATAT..... 1658  
228 SerLeuGluAsnLeuArgAlaLysTyrLeuAsnAlaGlyPheValArgPh 244  
: .....  
1659 .....CGAGTTGCTGTCAGGTATGCTTCT..... 1682  
244 eGluIleLysAspAlaLysLeuAsnIleAsnGluAspLysAsnArgIleP 261  
: .....  
1683 ....GTAACCCCGGATTCACCTCAAGCTTAATTTGGGGTAATTCATCCATT 1728  
261 he.....ValGluIleSerLeuHisGluGlyGln 271  
|| .....  
1729 TTTCCAATACAGTACCAGCTACAGCTACGTACATAGTAATACTACAAATCA 1778  
272 TyrArgPheGlyGln.....ThrGlnPheLeuG 281  
|||||.....  
1779 AGTGATTGTTGGTTATTTCGAAAGTGCATGCTTTTACATCTTCATTAGG 1828  
281 yAsnLeuThrTyrThrGln..... 287  
|||||.....  
1829 TAATATAGTAGTGTAGAAATTTTAGTGGGACTGCAGGAGTGATAATAG 1878

```

288 .....AlaGluLeuGluAlaLeuLeuLys 295
1879 ACAGATTGAATTTATCCAGTACTGCACACACTCGAGGCTGAATAAT 1928
296 PheLysAlaGluGluGlyPheSerGlnAlaMetLeuGluGlnThrAs 312
1929 CTGGAAGAGCGCAGAGCGGCTGAATCGCGTGTT...ACGCTACAAA 1975
312 nasIleSerThrLysPheGlyAspSerGlyTyrTyrAlaGlnIleA 329
1976 CCAACTAGGGCTAAAA..... 1991
329 rgProValThrArgIleAsnAspGluSerArgThrValAspValGluTyr 345
1992 .....ACAAATGTAACGGAT.....TAT 2009
346 TyrIleAspProValHisProValTyrValArgArgIleAsnPheThrG 362
2010 CATATTGATCAAGTG.....TC 2026
362 yAsnPheLysThr.....GlnAspGluValLeuArgArgGluMetArgG 377
2027 CAATTAGTTACGTATTATCGGATGAATTTGTCTGGATGAAAGCGAG 2076
377 InLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnLeuSerArgAla 393
2077 AATTGTCGAGAAGTCAACATCGGAGCGACTCAGTGATGACGC... 2123
394 ArgLeuMetArgThrGlyPhePheLysHisValThrValAspThrArgPr 410
2124 AATTACTCCAAGATTCAAAATTTCCAA..... 2150
410 ovalProAsnSerProAspGlnValAspValAsnPheValValGluGluG 427
2151 .....GACATTAAT.....AGGC 2163
427 InProSerGlySerSerThrIleAlaAlaGlyTyrSerGlnSerGlyGly 443
2164 AACAGAAA.....CGTGGTGGCGGAGGACACAGG 2195
444 ValThrPheGln.....PheAspValSerGlnAsnAsnPheMetGlyTh 458
2196 ATTACCATCCAAGAGGAGGTAGCGTATTTAAAGAAAAT..... 2234
458 rGlyLysHisValAsnAlaSerPheSerArgSerGluThrArgGluValT 475
2234 ..... 2234
475 yrSerLeuGlyMetThrAsnProTyrPheThrValAsnGlyValSerGln 491
2235 .....TACGTACACATATCAGGTACCTTGAT 2261
492 SerLeuSerGlyTyrTyrArgLysThrLysTyrAspAsn.....Ly 505
2262 GAGTGCTATCCAACATATTTGTATCAAAAATCGATGAATCAAAATANA 2311
505 sasIleSerAsnTyrValLeuAspSerTyrGlyLysLeuSerTyrG 522
2312 AGCCTTTACCGTTATCAATTAACAGGGTAT..... 2342
522 lyTyrProIleAspGluAsnGlnArgIleSerPheGlyLeuAsnAlaAsp 538
2343 .....ATCGAAGATAGTCAAGACTTAGAAATCTATTTAATTCGCTAC 2384
539 AsnThrLysLeuHisGlyGlyArgPheMetGlyIleSerAsnValLysG 555
2385 ANTCAAAACATGAACAGTAATATGTCGAGGTACGGGTTCTTATGGCC 2434
555 nLeuMetAlaAspGly.....GlyLysIleGlnValAspAsnAsnGlyI 570
2435 GCTTTTCAGCCCAAGTCCAATCGAAAGTGTGGAGAGCGGAATCGATCG 2484
570 leProAspPheLysHisAspTyrThrThrTyrAsnAlaIleLeuGlyTyr 586

```

```

2485 CGCCACAC.....CTTGAATGG 2501
587 AsnTyrSerSerLeuAspArgProValPheProThrGlnGlyMetSerH 603
2502 AAT...CTGACTTAGATTGCTGCTAGGATGGAGAAAAGTGTGCCCA 2548
603 s.....SerValAspLeuThrValGlyPheGlyAspLysT 615
2549 TCATTCCGATCATTTCTCTTAGACATGATGATAGGATGTACACACTTAA 2598
615 hrHisGlnLysValValTyrGlnGlyAsnIleTyrArgProPheIleLys 631
2599 ATGAGGACCTAGGTGTATGG.....GTGATCTTTAAG 2630
632 LysSerValLeuArgGlyTyrAlaLysLeuGlyTyrGlyAsnAsnLeuPr 648
2631 ATTAAGACGCAAGATGGCGACGCAAGACTAGGG.....AATCTAGA 2671
648 oPheTyrGlu 651
2672 GTTCTCGAA 2681

seq.name: /cqn2_6/ptodata/2/lna/5A_COMB.seq:us-07-828-788A-7

seq documentation block:
; Sequence 7, Application US/07828788A
; Patent No. 5273746
; GENERAL INFORMATION:
; APPLICANT: PAYNE, JEWEL M.
; APPLICANT: HICKLE, LESLIE A.
; TITLE OF INVENTION: NOVEL BACILLUS THURINGIENSIS ISOLATES
; TITLE OF INVENTION: ACTIVE AGAINST PHTHIRAPTERA PESTS
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID R. SALIWANCHIK
; STREET: 2421 N.W. 41ST STREET, SUITE A-1
; CITY: GAINESVILLE
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/828,788A
; FILING DATE: 19920129
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SALIWANCHIK, DAVID R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: MA75
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-372-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3531 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: BACILLUS THURINGIENSIS
; STRAIN: KURSTAKI
; INDIVIDUAL ISOLATE: PS81GG
; IMMEDIATE SOURCE:
; LIBRARY: LAMBDA GEM (TM) - 11 LIBRARY OF AUGUST SICK
; CLONE: 81GG

```

US-07-828-788A-7

alignment\_scores:  
 Quality: 130.00 Length: 720  
 Ratio: 0.406 Caps: 35  
 Percent Similarity: 44.444 Percent Identity: 19.583

alignment\_block:

US-09-701-711-2 x US-07-828-788A-7

Align seg 1/1 to: US-07-828-788A-7 from: 1 to: 3531

```

5 TyrPheLysGlyPheGlnValSerAlaMetThrMetAla..... 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
943 TATTGGTCAGGCGCATCAATAATAGCTTCTCCTGCTCGGTTTTCGGGGCC 992
18 .....ValMetMetValMetSerThrHisAlaGlnAlaAlaAspPheM 32
993 AGAATTACGTTTCGCGCTATATGGAACCATGGGAATGCAGCTCCA.... 1038
32 etAlaAsnAspIleThrIleThrGlyLeuGlnArgValThrIleGluSer 48
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1039 .....CAACAACGTATTGCTTCTCAACTA 1062
49 LeuGlnSerValLeu.....ProPh 55
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1063 GGTACGGCGCTATAGAACATTATCTCTACTCTTTTATAGAAGACCTTT 1112
55 eArgLeuGlyGlnValValSerGluAsnGlnLeuAla.....AspGlyV 70
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1113 TAAATATAGG.....ATAAATTAATCAACAACATATCTGTCTTGACGGGA 1156
70 aLysAlaLeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHis 86
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1157 CAGATTTCATTATGGAACCTCTCAAAATTCGCATCCGCTGTATACAGA 1206
87 GlnGluGlyArgIleIleTyrGlnValThrGluArgProLeuIleAlaGl 103
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1207 AAAAGCGGAACGGTA.....GATTCCGCTGATCA 1235
103 uileAsnPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluG 120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1236 AATACCACACAGATAACACAGCTCCACCTAGGCAAGGATTTAGTCATC 1285
120 lYLeuLysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThr 136
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1286 GATTAAAGCAT.....GTTTCAATGTTTCGTTACAGGCTCTAGTAGTAGT 1329
137 ValGlnMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTy 153
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1330 GTAAGTATAATAAGAGCTCTATGTTCTCTTGGATACATCGTAGTCTGA 1379
153 rTyrAsnThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgV 170
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1380 ATTTAAT...AATATAATTTCGATCGGATAGTATTACTCAAAATCCCTGCA 1426
170 aLysLeuAspMetThrPheAlaGluGlyLysProAlaArgValValasp 186
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1427 TGAAGGGAACCTTCTTTT.....AAT 1449
187 lIeAsnIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspVa 203
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1450 GGTTCGTGAATTTACAGACAGGATTTACTGCTGGGACTTAGTAGATT 1499
203 lIeuAlaIleLysAspAsnLysIleAsn.....ProL 214
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1500 AAATAGTAGTGGAAATAACATTCAAGATAGAGGGGTATATTGAAGTTCCAA 1549
214 eu.....SerLysAlaAspArgTyrThrGlnGluLysLeuValThr 227
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1550 TTCACCTTCCATCGACATCTACCAGATAT..... 1578

```

```

228 SerLeuGluAsnLeuArgAlaLysTyrLeuAsnAlaGlyPheValArgPh 244
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1579 .....CGAGTCGTACGGTATGCTTCT..... 1602
244 eGluIleLysAspAlaLysLeuAsnIleAsnGluAspLysAsnArgIleP 261
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1603 ...GTAACCCCGATTCACTCAACGTTAATTGGGGTAAATTCATCCATTT 1648
261 he.....ValGluIleSerLeuHisGluGlyGluGln 271
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1649 TTTCCAATACAGTACCAGCTACAGCTACGTCATTAGATAATCTACAATCA 1698
272 TyrArgPheGlyGln.....ThrGlnPheLeuGl 281
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1699 AGTGATTTTGGTTATTTGAAAGTGCCAATGCTTTTACATCTTTCATTAGG 1748
281 YAsnLeuThrTyrThrGln..... 287
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1749 TAATATAGTAGGTGTTAGAAATTTTAGTGGGACTGCAGGAGTGATAATAG 1798
288 .....AlaGluLeuGluAlaLeuLeuLys 295
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1799 ACAGATTTGAATTTATTCAGTTACTGCAACACTCGAGGCTGAATATAAT 1848
296 PheLysAlaGluGlyPheSerGlnAlaMetLeuGluGlnThrThrAs 312
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1849 CTGGAAGAGCGCAGAGGGGTGAATGCGCTGTTT...ACGCTACAAA 1895
312 nasIleSerThrLysPheGlyAspAspGlyTyrTyrTyrAlaGlnIleA 329
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1896 CCAACTAGGGCTAAAA..... 1911
329 rgProValThrArgIleAsnAspGluSerArgThrValaspValGluTyr 345
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1912 .....ACAAATGTACGGAT.....TAT 1929
346 TyrIleAspProValHisProValTyrValArgArgIleAsnPheThrGl 362
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1930 CATATTGATCAAGTG.....TC 1946
362 yAsnPheLysThr.....GluAspGluValLeuArgArgGluMetArgG 377
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1947 CAATTAGTACGTATTTATTCGGATGAATTTGCTGGATGAAAAGCGAG 1996
377 lNLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnLeuSerArgAla 393
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1997 AATTGTCGAGAAAGTCAACATCGCAAGCGCTCAGTGATGAACGC... 2043
394 ArgLeuMetArgThrGlyPhePheLysHisValThrValaspThrArgPr 410
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2044 AATTTACTCCAAGATTCAAAATTTCAAA..... 2070
410 oValProAsnSerProaspGlnValaspValAsnPheValValGluGluG 427
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2071 .....GACATTAAAT.....AGCG 2083
427 lNProSerGlySerSerThrIleAlaAlaGlyTyrSerGlnSerGlyGly 443
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2084 ACCAGAA.....CGTGGGTGGGCGCGGAAGTACAGGG 2115
444 ValThrPheGln.....PheAspValSerGlnAsnAsnPheMetGlyTh 458
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2116 ATTACCATCCAAGGAGGGGATGACGTATTTAAGAAAAAT..... 2154
458 rGlyLysHisValAsnAlaSerPheSerArgSerGluThrArgGluValT 475
2154 ..... 2154
475 yrSerLeuGlyMetThrAsnProTyrPheThrValAsnGlyValSerGln 491
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2155 .....TAGGTCACACTATCAGGTACCTTTGAT 2181
492 SerLeuSerGlyTyrTyrArgLysThrLysTyrAspAsn.....Ly 505

```

```

:::      |||      |||      |||      |||      |||
2182 GAGTGCATCAACATATTTGTATCAAAAATCGATGAATCAAAATAAA 2231
505 sAsnIleSerAsnTyrValLeuAspSerTyrGlySerLeuSerTyrG 522
2232 AGCCTTACCGTTATCAATAGAGGTAT..... 2262
522 lyTyrProIleAspGluAsnGlnArgIleSerPheGlyLeuAsnAlaAsp 538
2263 .....ATCGAAGATAGTCAAGACTTAGAAATCTATTTAAATTCGCTAC 2304
539 AsnThrIlyLeuHisGlyGlyArgPheMetGlyIleSerAsnValLysG 555
2305 AATGCAAAACATGAACAGTAATAGTGCAGGTACGGGTTCCTTATGGCC 2354
555 nLeuMetAlaAspGly.....GlyLysIleGlnValAspAsnAsnGlyI 570
2355 GCTTTCAGCCCAAGTCAATCGGAAAGTGTGGAGAGCCGAATCGATGCG 2404
570 leProAspPheLysHisAspTyrThrThrTyrAsnAlaIleLeuGlyT 586
2405 CGCCACAC.....CTTGAATGG 2421
587 AsnTyrSerSerLeuAspArgProValPheProThrGlnGlyMetSerH 603
2422 AAT...CCTGACTAGATTGTCGTAGGATGGAGAAAGTGGCCCA 2468
603 s.....SerValAspLeuThrValGlyPheGlyAspLysT 615
2469 TCATTCGCATCATTTCTCTTAGACATGTAGTGGATGTACAGACTTAA 2518
615 hrHisGlnLysValValTyrGlnGlyAsnIleTyrArgProPheIleLys 631
2519 ATGAGGACTAGGTATGG.....GTGATCTTTAAG 2550
632 LysSerValLeuArgGlyTyrAlaLysLeuGlyTyrGlyAsnAsnLeuPr 648
2551 ATTAAGACGAAGATGGCAGCAGACTAGGG.....AATCTAGA 2591
648 oPheTyrGlu 651
2592 GTTCTCGAA 2601

seq_name: /cgn2_6/ptodata/2/1na/5A_COMB.seq:US-07-920-085-1
seq_documentation_block:
; Sequence 1, Application US/07920085
; Patent No. 535661
; GENERAL INFORMATION:
; APPLICANT: PAYNE, JEWEL M.
; APPLICANT: SICK, AUGUST J.
; APPLICANT: THOMPSON, MARK
; TITLE OF INVENTION: NOVEL BACILLUS THURINGIENSIS ISOLATE DENOTED B.T.
; TITLE OF INVENTION: PS81GG, ACTIVE AGAINST LEPIDOPTERAN PESTS, AND A GENE
; TITLE OF INVENTION: ENCODING A LEPIDOPTERAN-ACTIVE TOXIN
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROMAN SALIWANCHIK
; STREET: 2421 N.W. 41st STREET, SUITE A-1
; CITY: GAINESVILLE
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/920,085
; FILING DATE: 19920727
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
```

1330 GTAAGTAATAAAGAGCTCCTATGTTCTTCTGGATACATCGTAGTGTCTGA 1379

153 TTrAsnThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgV 170

170 aLysLeuAspMetThrPheAlaGluGlyLysProAlaArgValValAsp 186

1427 TGAAGGGAAACTTCTCTTTT 1449

187 IleAsnIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspVa 203

1450 GGTCTGTAAATTCAGGACCGAGATTACTGGTGGGACTTAGTTAGTAT 1499

203 IleAlaIleLysAspAsnLysIleAsn 214

1500 AAATAGTAGTGGAAATACATTCAGATAGAGGCTATATTGAAGTTCCAA 1549

214 eu 227

1550 TTCACCTCCCATCGACATCTACCAGATAT 1578

228 SerLeuGluAsnLeuArgAlaLysTyrrLeuAsnAlaGlyPheValArgPh 244

1579 1602

244 eGluIleLysAspAlaLysLeuAsnIleAsnGluAspLysAsnArgIleP 261

1603 1648

261 he 271

1649 TTTCCAATACAGTACCAGCTACAGCTACGTCATAGATATATCTACATCA 1698

272 TyrArgPheGlyGln 281

1699 AGTGATTTTGGTTATTTTGAAGTGCCAAATGCTTTTACATCTCATTAGG 1748

281 yAsnLeuThrTyrrThrGln 287

1749 TAATATAGTAGTGTTAGAAATTTTATGGGACGTCGAGGAGTGATAATAG 1798

288 295

1799 ACAGATTGTAATTTATTCAGTTACTGCAACACTCGAGGCTGAATATAAT 1848

296 PheLysAlaGluGluGlyPheSerGlnAlaMetLeuGluGlnThrAs 312

1849 CTGGAAAGAGCGCAGGAAGCGGTGAATGCGGTGTT 1895

312 nAsnIleSerThrLysPheGlyAspAspGlyTyrrTyrrAlaGlnIleA 329

1896 CGAACTAGGGCTAANA 1911

329 rGProValThrArgIleAsnAspGluSerArgThrValAspValGluTyrr 345

1912 1929

346 TyrIleAspProValHisProValTyrrValArgArgIleAsnPheThrGl 362

1930 CATATTGTATCAAGTG 1946

362 yAsnPheLysThr 377

1947 CAATTTACTGTACGTATTTATCCGATGAATTTTGTCTGGATGAAGAAGCGAG 1996

377 InLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnLeuSerArgAla 393

1997 AATTGTCGGAAGACTCAACATCGGAAGCGCACTCAGTGATGAACGC 2043

394 ArgLeuMetArgThrGlyPhePheLysHisValThrValAspThrArgPr 410

2044 AATTACTTCCAAGATTCCAATTTTCCAA 2070

410 oValProAsnSerProAspGlnValAspValAsnPheValValGluGluG 427  
2071 .....GACATTAAT.....AGGC 2083  
427 InProSerGlySerSerThrIleAlaLaGlyTyrSerGlnSerGlyGly 443  
2084 AACCCAGAA.....CGTGGGTGGGCGCGGAAGTACAGGG 2115  
444 ValThrPheGln.....PheAspValSerGlnAsnAsnPheMetGlyTh 458  
2116 ATTACCATCCAAGGAGGGGATGACGTATTAAAGAAAT..... 2154  
458 rGlyLysHisValAsnAlaSerPheSerArgSerGluThrArgGluVal 475  
2154 ..... 2154  
475 yrSerLeuGlyMetThrAsnProTyrPheThrValAsnGlyValSerGln 491  
2155 .....TACGTCAACATCATCAGTACCTTTGAT 2181  
492 SerLeuSerGlyTyrTyrArgLysThrLysTyrAspAsn.....Ly 505  
2182 GAGTGCTATCCACATATTGCTATCAAAAATCATCATCAAAATTTAA 2231  
505 sAsnIleSerAsnTyrValLeuAspSerTyrGlyGlySerLeuSerTyrG 522  
2232 AGCCTTTACCCTTATCAATTAAAGGGTAT..... 2262  
522 LyTyrProIleAspGlnAsnGlnArgIleSerPheGlyLeuAsnAlaAsp 538  
2263 .....ATCGAAGATAGTCACAGACTTAGAAATCTATTAAATTCGCTAC 2304  
539 AsnThrLysLeuHisGlyGlyArgPheMetGlyIleSerAsnValLysG 555  
2305 ATGCGAAACATGAACAGTAATGTCGAGGTACGGGTTCCTTATGGCC 2354  
555 nLeuMetAlaAspGly.....GlyLysIleGlnValAspAsnGlyI 570  
2355 GCTTTACGCCCAAGTCCAAATCGGAAAGTGTGGAGAGCCGAATCGATGC 2404  
570 leProAspPheLysHisAspTyrThrThrTyrAsnAlaIleLeuGlyTyr 586  
2405 CGCCACAC.....CTGAATGG 2421  
587 AsnTyrSerSerLeuAspArgProValPheProThrGlnGlyMetSerHi 603  
2422 AAT...CTGACTAGATTGTCGTAGGATGGAGAAAGTGCCTCA 2468  
603 s.....SerValAspLeuThrValGlyPheGlyAspLysT 615  
2469 TCATTCCGCATCATTTCTCCTTAGACATTGATAGGATGTACAGACTAA 2518  
615 hrHisGlnLysValValTyrGlnGlyAsnIleTyrArgProPheIleLys 631  
2519 ATGAGGACCTAGGTGTATGG.....GTGACTTTTAAG 2550  
632 LysSerValLeuArgGlyTyrAlaLysLeuGlyTyrGlyAsnAlaLeuPr 648  
2551 ATTAGACGCAAGATGGCGACGAAGCTAGGG.....AATCTAGA 2591  
648 oPheTyrGlu 651  
2592 GTTTCCTCGAA 2601

```

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq: PCT-US92-11337-7

seq_documentation_block:
; Sequence 7, Application PC/TUS92111337
; GENERAL INFORMATION:
; APPLICANT: PAYNE, JEWEL M.
; APPLICANT: HICKLE, LESLIE A.
; TITLE OF INVENTION: NOVEL BACILLUS THURINGIENSIS ISOLATES

```

```

; TITLE OF INVENTION: ACTIVE AGAINST PHTHIRAPTERA PESTS
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID R. SALIWANCHIK
; STREET: 2421 N.W. 41st STREET, SUITE A-1
; CITY: GAINESVILLE
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/11337
; FILING DATE: 19921231
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 97/828,788
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: SALIWANCHIK, DAVID R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: MAY5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3531 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: BACILLUS THURINGIENSIS
; STRAIN: KURSTAKI
; INDIVIDUAL ISOLATE: PS81GG
; IMMEDIATE SOURCE:
; LIBRARY: LAMBDA GEM (TM) - 11 LIBRARY OF AUGUST SICK
; CLONE: 81GG
; PCT-US92-11337-7

alignment_scores
  Quality: 130.00 Length: 720
  Ratio: 0.406 Gaps: 35
  Percent Similarity: 44.444 Percent Identity: 19.583

alignment_block
US-09-701-711-2 x PCT-US92-11337-7 ..
Align seg 1/1 to: PCT-US92-11337-7 from: 1 to: 3531

5 TyrPheLysGlyPheGlnValSerAlaMetThrMetAla.....17
|||||
943 TATTGGTCAGGCATCAATATATGCTTCTCCTGTCGGTTTTCGGGGCC 992
18 .....ValMetMetValSerThrHisAlaGlnAlaAlaAspPheM 32
|||||
993 AGAATTCAGTTCCGCTATATGGAACCATGGGAATGCAGCTCCA....1038
32 etAlaAsnAspIleThrIleThrGlyLeuGlnArgValThrIleGluSer 48
|||||
1039 .....CAACACGTAATTGTGCTCAACTA 1062
49 LeuGlnSerValLeu.....ProPh 55
|||||
1063 GGTGAGCGGTGTATAGAACATTATCTCTACTTTTATAGAACCTTT 1112

```

```

55 eArgLeuGlyGlnValValSerGluAsnGlnLeuAla.....AspGlyV 70
|||||
1113 TAATATAGG.....ATAAATAATCAACACTACTCTGTTCTTACGGGA 1156

70 allysAlaLeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHis 86
|||||
1157 CAGAAATTGCTTATGGAACCTCTCAAAATTGGCCATCCGCTGTATACAGA 1206

87 GlnGluGlyArgIleIleTyrGlnValThrGluArgProLeuIleAlaGl 103
|||||
1207 AAAAGCGGAACGGTA.....GATTGCTGGATGA 1235

103 uileAsnPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluG 120
|||||
1236 AATACCACACAGAAATCAACAGCTGCCACCTAGGCAAGGATTTAGTCATC 1285

120 lYleuLysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThr 136
|||||
1286 GATTAAAGCCAT.....GTTTCAATGTTTCGTTAGGCTCTAGTAGTAGT 1329

137 ValGlnMetIleGluThrGluThrAsnGlnTyrIleSerGlnGlyTy 153
|||||
1330 GTAAGTATAATAAGAGCTCTATGTTCTCTGGATACATCTAGTAGTCTGA 1379

153 rTyAsnThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgV 170
|||||
1380 ATTTAAT...AATAAATTGCATCGGATAGTATTACTCAATCCCTGCAG 1426

170 allysLeuAspMetThrPheAlaGluGlyLysProAlaArgValValasp 186
|||||
1427 TGAAGGGAACCTTCTTT.....AAT 1449

187 lIeAsnIlelleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspVa 203
|||||
1450 GGTCTCTGTAATTCAGACAGGATTTACTGTGGGACTTAGTTAGATT 1499

203 lIeAlaIleLysAspAsnLysIleAsn.....ProL 214
|||||
1500 AAATAGTAGTGAAATATACATTCAGAAATAGAGGGTATATTGAAGTTCCA 1549

214 eu.....SerLysAlaAspArgTyrThrGlnGluLysLeuValThr 227
|||||
1550 TTCACCTTCCCATCGACATCTACCAGATAT..... 1578

228 SerLeuGluAsnLeuArgAlaLysTyrLeuAsnAlaGlyPheValArgPh 244
|||||
1579 .....CGAGTTCGTGTCAGGTATGCTTCT..... 1602

244 eGluIleLysAspAlaLysLeuAsnIleAsnGluAspLysAsnArgIleP 261
|||||
1603 ....GTAACCCCGATTCACCTCAACGTTAATTGGGGTAATTCATCATTT 1648

261 he.....ValGluIleSerLeuHisGluGlyGluGln 271
|||||
1649 TTTCCAATACAGTACCAGCTACAGCTACGCTACGCTATAGATAATCTACAATCA 1698

272 TyrArgPheGlyGln.....ThrGlnPheLeuGl 281
|||||
1699 AGTGATTTTGGTTATTTTGAAGTGCAATGCTTTTACATCTCATAGG 1748

281 yAsnLeuThrTyrThrGln..... 287
|||||
1749 TAATATAGTAGGTGTAGAAATTTTAGTGGGACTGCAGGAGTGATAATAG 1798

288 .....AlaGluLeuGluAlaLeuLeuLys 295
|||||
1799 ACAGATTTGAAATTTATTCAGTTACTGCAACACTCGAGGCTGAATATAAT 1848

296 PheLysAlaGluGlyPheSerGlnAlaMetLeuGlnThrThrAs 312
|||||
1849 CTGGAAGACGCGAGAGGGGTGAATGCGCTGTTT...ACGTCACAAA 1895

312 nAsnIleSerThrLysPheGlyAspAspGlyTyrTyrTyrAlaGlnIleA 329

```



```

1896 CCAACTAGGCTAAA..... 1911
329 rgProValThrArgIleAsnAspGluSerArgThrValaspValGluTyr 345
1912 .....ACAAATGTAACGGAT.....TAT 1929
346 TyrIleAspProValHisProValTyrValArgArgIleAsnPheThrGlu 362
1930 CATATTGATCAAGTG.....TC 1946
362 yAsnPheLysThr.....GlnAspGluValLeuArgGluMetArg 377
1947 CAATTAGTATACGTATTATTATCGGATGAATTTCTCTGGATGAAAAGCGAG 1996
377 lnLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnLeuSerArgAla 393
1997 AATTGTCCGAGAAAGTAAACATGCGAAGCGACTCAGTGATGAACGC... 2043
394 ArgLeuMetArgThrGlyPhePheLysHisValThrValAspThrArgPr 410
2044 AATTACTCAAGATTCAAATTTCAAA..... 2070
410 oValProAsnSerProAspGlnValaspValAsnPheValValGluGluG 427
2071 .....GACATTAAT.....AGGC 2083
427 lnProSerGlySerSerThrIleAlaAlaGlyTyrSerGlnSerGlyGly 443
2084 ACCCAGAA.....CGTGGGTGGGGCGGAGATACAGGG 2115
444 ValThrPheGln.....PheAspValSerGlnAsnAsnPheMetGlyTh 458
2116 ATTACCATCAAGAGGGGATGACGTATTTAAGAAAAAT..... 2154
458 rGlyLysHisValAsnAlaSerPheSerArgSerGluThrArgGluValT 475
2154 ..... 2154
475 yrSerLeuGlyMetThrAsnProTyrPheThrValAsnGlyValSerGln 491
2155 .....TAGTCACACTATCAGGTACCTTTGAT 2181
492 SerLeuSerGlyTyrTyrArgLysThrLysTyrAspAsn.....Ly 505
2182 GAGTGCTATCCAACATATTGTATCAAAAAATCGATGAATCAAAATTAA 2231
505 sAsnIleSerAsnTyrValLeuAspSerTyrGlySerLeuSerTyrG 522
2232 AGCCTTTACCCGTTATCAATTTAAGAGGGTAT..... 2262
522 lyTyrProIleAspGluAsnGlnArgIleSerPheGlyLeuAsnAlaAsp 538
2263 .....ATCGAAGATAGTCAAGACTTAGAATCTATTAAATCGCTAC 2304
539 AsnThrLysLeuHisGlyGlyArgPheMetGlyIleSerAsnValLysG 555
2305 AATGCAAAACATGAACAGTAATGTCCAGGTACGGGTTCTCTATGGCC 2354
555 nLeuMetAlaAspGly.....GlyLysIleGlnValAspAsnAsnGlyI 570
2355 GCTTTACGCCCAAGTCCAAATCGGAAAGTGTGGAGAGCCCGAATCGATCG 2404
570 leProAspPheLysHisAspTyrThrThrTyrAsnAlaIleLeuGlyT 586
2405 CGCCACAC.....CTTGAATGG 2421
587 AsnTyrSerSerLeuAspArgProValPheProThrGlnGlyMetSerHi 603
2422 AAT...CCTGACTTAGATTGTCGTAGGGATGGAGAAAAAGTGCCTCA 2468
603 s.....SerValAspLeuThrValGlyPheGlyAspLysT 615

```

---

```

2469 TCATTGCGCATCAATTCCTCCTTAGACATTGATGTAGGATGTACAGACTTAA 2518
615 hrHisGlnLysValValTyrGlnGlyAsnIleTyrArgProPheIleLys 631
2519 ATGAGGACCTAGGTGTATGG.....GTGATCTTTAAG 2550
632 LysSerValLeuArgGlyTyrAlaLysLeuGlyTyrGlyAsnAsnLeuPr 648
2551 ATTAAGACGCAAGATCGGCACGCAAGACTAGGG.....AATCTAGA 2591
648 oPheTyrGlu 651
2592 GTTCTCGAA 2601

```

